Journal of Applied Psychology

Edited by Donald G. Paterson, University of Minnesota

Consulting Editors

George K. Bennett, *Psychological Corporation*
Harold E. Burtt, *Ohio State University*
Allen L. Edwards, *University of Washington*
Clifford E. Jurgensen, *Minneapolis Gas Co.*
Irving Lorge, *T. C. Columbia University*
Quinn McNemar, *Stanford University*
Alexander Mintz, *City College of New York*
James P. Porter, *Claverack, New York*
Harold F. Rothe, *Fairbanks, Morse and Co., Beloit, Wis.*
Julian B. Rotter, *Ohio State University*
Edward K. Strong, Jr., *Stanford University*
Donald E. Super, *T. C. Columbia University*
Morris S. Viteles, *University of Pennsylvania*
Alfred C. Welch, *Knox-Reeves, Minneapolis*

Table of Contents

Socio-Psychological Factors in Industrial Morale: II: R. E. Bernberg 249
Predicting Success in Dental School: W. L. Layton 251
Prediction and Practice Tests at the College Level: S. B. Anderson 256
Development of a Short Test to Predict a Complex Aggregate Score: H. Tomlinson and J. T. Preston 260
The Classification of Occupations by Means of Kuder Interest Profiles: I. The Development of Interest Groups: J. L. Holland, A. H. Krause, M. E. Nixon, and M. F. Trembath 263
The Validity of the Mooney Problem Check List: C. J. McIntyre 270
A Comparison of Manual and College Norms for the MMPI: F. T. Tyler and J. U. Michaelis .. 273
Socio-economic Status and Culturally-weighted Test Scores of Negro Subjects: F. C. J. McGurk .. 276
The Relationship Between Rater Characteristics and Validity of Ratings: D. E. Schneider and A. G. Bayroff 278
The Actuality Measure in the Study of Public Opinion: P. R. Hofstaetter 281
A Biasing Factor in Essay Response Frequency: E. K. Taylor and D. E. Schneider 288
Rating Patterns for Maximizing Competition and Minimizing Number of Comparative Judgments Necessary for Each Rater: R. H. Simpson 290
Some Primary Ratable Characteristics of Instructional Films: P. Ash and T. R. Hobaugh 293
The Relation of Light Intensity to Accuracy of Depth Perception: A. S. Edwards 300
Instrument Reading III: Check Reading of Instrument Groups: W. J. White, M. J. Warrick, and W. F. Grether 302
Dimensional Analysis of Motion: VI. The Component Movements of Assembly Motions: R. Smader and K. U. Smith 308
The F Minus K Index on the MMPI: A. G. MacLean, A. T. Tait, and C. D. Catterall 315
Psychological and Personal History Data Related to Accident Records of Commercial Truck Drivers: J. W. Parker, Jr. 317
Applied Psychology in Action:
A New Management Tool for Top Executives 321
Background of an Industrial Psychologist 321
Noise and Absenteeism 322
Book Reviews 323
New Books, Monographs, and Pamphlets 328

American Psychological Association

Vol. 37, No. 4 **August, 1953**

Journal of Applied Psychology

Published Bi-monthly by the American Psychological Association, Inc.
Prince and Lemon Sts., Lancaster, Pa.

Annual subscription, $6.00; single copies, $1.25

Subscriptions and business communications should be sent to
American Psychological Association
1333 Sixteenth Street N.W.
Washington 6, D. C.

Articles for publication and books for review should be sent to the Editor
Professor Donald G. Paterson, Department of Psychology
University of Minnesota, Minneapolis 14, Minnesota

This Journal gives prompt consideration to manuscripts reporting *original investigations* in any field of applied psychology except clinical and consulting psychology. A descriptive or theoretical article is occasionally accepted if it deals in a distinctive manner with a problem of applied psychology. The policy is, however, to favor papers dealing with quantitative investigations of direct value to psychologists working in the following fields: Vocational diagnosis and occupational guidance; educational diagnosis, prediction and guidance at the secondary school level and higher; personnel selection, training, placement, transfer and promotion in business, industry and government service including the armed forces; supervisory training in business, industry and government; bio-mechanics or design of machines to fit the human operator; illumination, ventilation and fatigue in industry; job analysis, description, classification and evaluation; measurement of morale of executives, supervisors, or employees; surveys of opinion on social or political issues, such as those conducted by The Psychological Corporation; psychological problems in market research and in advertising.

Articles may be under 500 words. The maximum is 12,000 words, the average in the neighborhood of 4,000 words. To reduce lag of publication, adherence to the rule of "brevity consistent with clarity" is encouraged.

A lapse of six to twelve months occurs between acceptance of an article and its publication, the lag varying with the rate at which manuscripts are submitted. If, however, an author is prepared to defray the costs of printing the necessary extra pages, he may arrange for earlier publication without thereby postponing the appearance of manuscripts by other contributors. This enables the management to provide space in addition to the scheduled 64 pages per issue. "Early publication" is thus a direct contribution to the subscribers. By cutting down lag in publication, it also benefits those authors whose articles are published in regular turn.

Tables, footnotes and references as well as text of manuscripts should be typed double-spaced throughout. Authors should adhere to the conventions described in the "Publication Manual of the American Psychological Association," *Psychol. Bull.*, 1952, 49, No. 4, Part 2. A copy of the Manual will be loaned to any prospective contributor who does not find it in his library.

Entered as second-class matter, August 19, 1943, at the post office at Lancaster, Pa., under the act of March 3, 1879

Acceptance for mailing at the special rate of postage provided for in paragraph (d-2), Section 34.40, P. L. & R. of 1948, authorized October 10, 1947

Copyright, 1953, by The American Psychological Association, Inc.

Journal of Applied Psychology

VOL. 37, NO. 4 AUGUST, 1953

Socio-Psychological Factors in Industrial Morale: II

Raymond E. Bernberg

Los Angeles State College

Previously reported research (1) compared the predictive ability of different tests of morale for performance indicators in the work situation. The tests of morale were based upon current concepts; viz.: group morale (GM); employee attitude toward the company (CM) (i.e., acceptance of the formal organization by its members); rating of the supervisors by the workers (S); and self-rating of morale by the workers (SM). Performance indicators used were absences, tardiness, short time absences, medical-aid unit visits, and merit rating. The results indicated a general failing of all tests of morale to have much predictive value for any of the performance indicators.[1]

An interesting result showed up when SM (a thermometer scale requesting the worker to check along a 0 to 100-degree parameter with verbal referents the degree to which he agreed with the proposition that his work group had high morale exemplified by their working together as a well organized team) was taken as a criterion and a determination made of the predictive value the other three tests of morale had for it.

Table 1

Intercorrelations of the Tests of Morale

	GM	CM	S	SM
GM	—	.77	.51	.67
CM	.77	—	.48	.47
S	.51	.48	—	.49
SM	.67	.47	.49	—

Table 2

Beta Weights and Multiple R with SM as Criterion

	1 (GM)	2 (CM)	3 (S)
β	.71	−.16	.16

$R_{0.123} = .69$; $\sigma R_{0.123} = .02$.

Table 1 contains the Pearson product moment intercorrelations of the four measures based upon 890 cases. The results of the multiple regression analysis are in Table 2. It appears that GM with a correlation of plus .67 with SM has contributed almost the entirety of the multiple R of plus .69.

This result gives much validity to the concept of morale as a group phenomenon as measured by the group morale test. This, of course, is based upon the assumption that the collective opinions of the workers themselves are an adequate criterion for appraising morale in the work situation.

The group morale test (2) is a projective type paper and pencil test using the direction of perception technique of attitude measurement. It is based upon content derived from six determiners discussed by Krech and Crutchfield (4). They are: 1) positive goals; 2) satisfaction of accessory needs; 3) sense of advance toward goals; 4) level of aspiration and achievement; 5) time perspective; and 6) feelings of identification, solidarity, and involvement. There are 34 items in the test, all equally weighted.

The question which presented itself next was: If this test with its total content has

[1] This conclusion is in conflict with the multiple correlation of .71 between six objective indices of efficiency and morale scores. See Giese, J. G., and Ruter, H. W. An objective analysis of morale. *J. appl. Psychol.*, 1949, 33, 421–428.—*Editor.*

such a high relationship with the collective opinion of workers concerning their cognition of morale in their work group, which items with their specific content comprise the maximal possible prediction for the test as a whole? This is an analytical question which in essence asks: What do the workers mean by morale? This of course concerns only the relationship of the content of the items of the test and the criterion.

Gengerelli (3) has developed a technique of analysis whereby one may reduce a large battery of tests in their relation to a criterion, to a number of sub-tests which provide the maximal prediction of the criterion. This method describes the whole battery in terms of the smaller sub-set and provides a regression equation to express the total correlation

Table 3

Intercorrelations of the Four "Factors" (Items) and SM

	Item				
Item	SM	3	9	11	18
SM	—	.17	.14	.32	.82
3	.17	—	.16	.00	−.14
9	.14	.16	—	.12	−.34
11	.32	.00	.12	—	.17
18	.82	−.14	−.34	.17	—

matrix. He states that this method yields the sub-tests as empirical tests which might be considered as "factors" of a sort. This provides an immediate practical solution to answer the analytical question posed above.

The SM measure was taken as the criterion and the 34 items of the GM test as the battery of tests for the analysis. Because of the large number of cases, 100 of the 890 were selected randomly for determining the intercorrelations between the items (tetrachoric *r*) and the correlations between the items and the criterion (bi-serial *r*).

The results of the analysis produced 4 factors which as a sub-set have a multiple R plus .96 with the criterion. Table 3 shows the intercorrelations between these items and with the criterion. Table 4 indicates their beta weights.

Table 4

Beta Weights and Multiple R with SM as Criterion

	1 (Item 3)	2 (Item 9)	3 (Item 11)	4 (Item 18)
β	.218	.419	.130	.820

$R_{0.1234} = .96$; $\sigma R_{0.1234} = .09$.

These four "factors" are the following items of the test: Item 3. "Scientific studies show that in groups such as yours, if you wanted to hold a big party, picnic or other type of friendly gathering, you would not care to invite: (a) *60% of your work group;* (b) *35% of your work group.* Item 9. "Statistics show continually that the increase in group production is a result of: (a) *group effort toward step-up;* (b) *individual effort.* Item 11. "Recent industrial studies have shown that the following percentage of workers in a group such as yours gives a good amount of attention to ways of getting ahead: (a) *60%;* (b) *90%.* Item 18. "A recent poll of workers in groups such as yours found that workers got a lot of satisfaction from working together: (a) *infrequently;* (b) *frequently.*"

The possible importance of these findings need not lie in the measurement of morale per se. It appears that it would be wise in developing and controlling work groups to consider the following factors: (1) satisfaction of men from working together; (2) increase in production as a result of group effort; (3) intimacy of workers with each other beyond as well as in the work surroundings; and (4) the individual level of aspiration in getting ahead.

Received October 3, 1952.

References

1. Bernberg, R. E. Socio-psychological factors in industrial morale: I. The prediction of specific indicators. *J. soc. Psychol.*, 1952, **36**, 73–81.
2. Bernberg, R. E. The direction of perception technique of attitude measurement. *Int. J. Opin. Attit. Res.*, 1951, **5**, 397–406.
3. Krech, D. and Crutchfield, R. S. *Theory and problems of social psychology.* New York: McGraw-Hill, 1948.
4. Gengerelli, J. A. A method of analysis in which the factors are empirical tests. *J. Psychol.*, 1952, **33**, 159–174.

THE JOURNAL OF APPLIED PSYCHOLOGY
Vol. 37, No. 4, 1953

Predicting Success in Dental School

Wilbur L. Layton

Student Counseling Bureau, University of Minnesota

During the past several years, a trend in the field of testing has been indicated by the establishment of nation-wide testing programs in which applicants for admission to certain schools, mainly professional schools, are tested in centralized programs. All participating colleges or schools then use the resulting test scores in the process of selecting and admitting students.

However, a test battery which is valid as a selection device at one institution may not be valid at another institution. Rarely do the administrators of the national testing programs publish validity data pertaining to specific institutions. Hence most institutions participating in a national program are doing so blindly unless they evaluate the test battery in their own situation.

Since the fall of 1946, the School of Dentistry at the University of Minnesota has participated in the testing program sponsored by the Council on Dental Education of the American Dental Association by testing freshman classes at the beginning of the school year. During the first five years of this testing program, an attempt was made to determine the relationship between various tests administered in the program and success in Dental School. In 1951 the Council on Dental Education decided that the test battery currently used was adequate and should be used by dental schools to select students for their freshman classes. However, little evidence was available locally to indicate the usefulness of the battery for selection purposes.

Consequently, a study was designed that would correlate pre-dental grades and the test data available through the national program with grades earned by dental school freshmen who entered the University of Minnesota Dental School in the years 1946 through 1949. First year grades in the dental school were the primary criteria for each of the four classes. For the freshman class entering in 1946 four year grades were also available as criteria. Separate grades in freshman courses in physiological chemistry, anatomy and prosthetics were also used as criteria for all four classes studied. These data were punched into Hollerith cards to facilitate analysis.

Table 1 presents the means and standard deviations for the data which were available for the classes entering in the various years 1946 through 1949. Table 1 indicates that the tests included in the battery varied from year to year. The battery was stabilized in 1949. This battery is the one currently (1952) being used in the national program.

The *ACE* is the American Council on Education psychological examination. It is one of the most widely used scholastic aptitude tests. It provides three scores,—a linguistic or *L* score, a quantitative or *Q* score, and a total score. Only the total score was used in this investigation.

The Survey of Object Visualization is similar in content to the Revised Minnesota Paper Form Board. *The Survey of Natural Sciences* is a 90-item test measuring facts and applications of principles in biology, chemistry and physics. The total raw score was used in the analyses.

The Carving Dexterity test is a chalk carving test developed by the Council on Dental Education. The examinee uses a knife, ruler and pencil to carve a piece of chalk which measures approximately 3¼" in length and ⅝" in diameter to correspond to an illustrated figure. The carvings are graded by judges on the basis of objectives such as: "flatness of surfaces," "clean-cutness of angles," "symmetry" and "accuracy of reproduction." The test supposedly measures finger-knife dexterity as well as spatial visualization.

GED No. 3 is the USAFI test of General Educational Development test three, "Interpretation of Reading Materials in the Natural Sciences." It measures speed of reading and comprehension of passages in the natural sciences. In 1947, 1948 and 1949 only the

Table 1

Means and Standard Deviations of Criteria and Predictive Variables for Groups Entering the Dental School in 1946, 1947, 1948 and 1949

	1946 N = 81		1947 N = 80		1948 N = 84		1949 N = 88	
	M	S.D.	M	S.D.	M	S.D.	M	S.D.
ACE	131.9	20.9	136.3	15.9	127.8	18.5	134.1	17.5
Carving	13.3	3.1	11.4	2.6	9.9	2.3	11.8	2.9
Survey of Obj. Vis.	29.2	7.4	28.9	6.4	32.1	6.0	30.9	7.5
Survey of Sciences	*		55.4	8.0	55.4	6.9	55.3	9.0
GED No. 3	55.5	12.1	34.0	5.3	34.2	5.2	34.7	5.2
GED No. 1 A	16.6	4.0	16.0	4.0	15.6	3.6	*	
GED No. 1 B	82.2	11.4	83.2	11.1	83.1	9.4	*	
Mich. Vocabulary								
Phys. Science	20.1	3.8	*		*		*	
Biol. Science	18.4	4.9	*		*		*	
Peterson Word Dex.	35.8	8.6	*		*		*	
Pre-dent. HPR**	1.4	.4	1.8	.4	1.8	.4	1.7	.5
Frosh HPR**	1.5	.4	1.7	.4	1.7	.4	1.4	.5
Four year HPR**	1.6	.3						
Anatomy HPR**	1.6	.5	1.5	.5	1.5	.5	1.4	.7
Phys. Chem. HPR**	1.5	.6	1.5	.8	1.8	.7	1.3	.8
Prosthetics HPR**	1.9	.4	1.9	.3	1.7	.4	1.8	.5

* Not given.

** HPR = $\frac{\text{honor points}}{\text{number of credits}}$ (A = 3; B = 2; C = 1; D, F = 0 honor points).

first 45 out of the total 90 items were given.

GED No. 1 is the USAFI test of General Educational Development test one, "Correctness and Effectiveness of Expression." It consists of a section on spelling and a section in which corrections are to be made in punctuation, words and phrases in a connected text. In Table 1, *A* is the score on the spelling section of the test; *B* is the score on the total test.

Table 2

Coefficients of Various Predictive Indices with Freshman and Cumulative (4 year) Honor Point Ratios in the Dental School for the 81 Students in the Class Entering in 1946

Predictor	Freshman Grades	Four Year Grades
Freshman HPR	—	.83
ACE Total	.15	.09
GED Reading	.23	.10
GED No. 1 A	.40	.27
GED No. 1 B	.40	.24
Word Dexterity	.29	.25
Mich. Vocab. (Phys. Science)	.29	.12
Mich. Vocab. (Biol. Science)	.37	.19
Survey Obj. Vis.	.14	.49
Carving	.22	.31
Pre-dental HPR	.40	.11

The Michigan Vocabulary is the Michigan Vocabulary Profile test. The two parts "physical science vocabulary" and "biological science vocabulary" were included in the test battery in 1946.

The Peterson Word Dexterity test measures specifically the extent to which the student knows the meaning of certain suffixes and prefixes and measures "dexterity" at manipulating word parts and word meanings.

Coefficients of correlation between the predictive indices and total first year grades and grades in the three first year courses in the Dental School were computed for all four groups. Four year grades for the group entering in 1946 were correlated with the predictive indices also.

A multiple correlation and a regression equation were computed for the 1949 group.

Table 3

Correlation Coefficients of the Various Predictive Indices with Freshman Honor Point Ratio in the Dental School for Classes Entering in 1946 Through 1949

Criterion: Freshman HPR	ACE (Total)	Survey Obj. Vis.	Carving	Survey of Sciences†	GED No. 3*	Predent. HPR	No. of Cases
1946	.15	.14	.22	—	.23	.40	81
1947	.13	.38	.20	.43	.37	.27	80
1948	.11	.29	.21	.17	−.01	.40	84
1949	.34	.29	.18	.43	.34	.42	88

† Not given in 1946.
* Only ½ of test given to 1947, 1948 and 1949 groups.

This group was selected because it was given the tests currently being given in the battery. These tests presumably are the most valid on a nationwide basis.

Table 2 presents the coefficients of correlation between the predictive indices and freshman and four year grades for the class entering in 1946. With an N of 81 the standard error of r if the true r is zero is .11.

Freshman grades correlated rather highly ($r = .83$) with total four year grades. It is interesting to note that the correlations of the predictive variables with four year grades are smaller than those with freshman grades alone. However, the *Survey of Object Visualization* and *Carving Dexterity tests* are more highly related to four year grades. This may be due to the small number of courses requiring these skills that are given during the first two years of the curriculum. The last two years of the curriculum are heavily loaded with practicum and clinic courses which require these skills of the students.

Table 3 presents the correlation coefficients for various predictive indices for each of four classes with freshman honor point ratio in the dental school. Tests given in Table 3 are the tests which are currently used in the National Testing Program plus the pre-dental honor point ratio.

In three of the four years (1947 being the exception) the pre-dental honor point ratio was the best predictor of freshman honor point ratios in the dental school. The group entering in 1947 was heavily loaded with veteran students who had taken their pre-dental work before the war. The difference in motivation for these students in their pre-war school work and their post-war school work may account for the relatively lower correlation in the 1947 group. The findings of Hansen and Paterson (1) that there is a "striking increase in post-war scholastic achievement as compared with pre-war scholastic achievement of the same students" (all veterans) tend to support this interpretation. It is interesting to note that the ACE score did not correlate highly with freshman honor point ratio in the first three classes studied, but did for the group entering in 1949. The *Survey of Sciences test* cor-

Table 4

Intercorrelations of the Variables for the 88 Students in the Class Entering the Dental School in 1949

	ACE Total	Survey of Object Vis.	Carving Dexterity	Survey of Sciences	GED No. 3	Pre-dental HPR
Fresh. HPR	.34	.29	.18	.43	.34	.42
ACE Total		.20	.17	.37	.43	.41
Survey of Object Vis.			.43	.29	.27	.02
Carving Dexterity				.15	.17	.11
Survey of Sciences					.46	.28
GED No. 3						.24

Table 5

Correlations of Honor Point Ratios in Oral Anatomy 50–51–52, Prosthetics 50–51–52, Physiology 58–59 with Five Predictive Measures for Four Classes of Dental School Freshmen

	1946† N = 81	1947 N = 86	1948 N = 90	1949 N = 85
a. Oral anatomy 50–51–52 HPR with:				
1. Pre-dental HPR	.37	.26	.26	.41
2. ACE Total	.12	.08	.10	.37
3. Survey of Obj. Vis.	.35	.32	.37	.34
4. Carving	.42	.30	.31	.24
5. Survey of Sciences	—	.18	.17	.38
6. GED No. 3	.08	.11	−.04	.27
b. Prosthetics 50–51–52 HPR with:				
1. Pre-dental HPR	.28	.11	.05	.26
2. ACE Total	.00	.20	.09	.43
3. Survey of Obj. Vis.	.21	.29	.13	.16
4. Carving	.29	.11	.01	.01
5. Survey of Sciences	—	.49	.14	.51
6. GED No. 3	−.15	.09	−.14	.04
c. Physiology 58–59 HPR with:				
1. Pre-dental HPR	.22	.25	.41	.34
2. ACE Total	.05	.04	−.06	.14
3. Survey of Obj. Vis.	.07	.11	.35	.19
4. Carving	.18	.22	.36	.33
5. Survey of Sciences	—	.05	.14	.10
6. GED No. 3	.14	.50	.03	.39

† Survey of Sciences Test was not given to the Freshman Class in 1946.

relates higher with first year honor point ratio for the 1947 class and for the 1949 class than for the 1948 class.

The fluctuation of the coefficients of correlation from year to year may be due to the differences from year to year in the means and standard deviations for the variables and hence differences in the composition of the classes. Changes in the curriculum could also account for the variability of the correlation coefficients. Also, because of the small number of cases studied in the various classes, this variation may be only random variation.

The class entering in 1949 was used for the computation of intercorrelations of the several variables, a coefficient of multiple correlation and a regression equation.

Table 4 presents the intercorrelation of the variables studied.

The regression equation and coefficient of multiple correlation were computed by the Doolittle method. Of the six predictor variables only two, *Survey of Sciences* and pre-dental *HPR*, yielded significant Beta coefficients. The coefficient of multiple correlation obtained was .54.

The regression equation is:

$$\bar{y} = .02\ X_1 + .39\ X_2 - .29.$$

X_1 = Survey of Sciences total raw score;
X_2 = Pre-dental *HPR;*
$\bar{y}$ = Predicted freshman year honor point ratio.

Table 5 presents the coefficients of correlation between five predictive indices and honor point ratios in oral anatomy, prosthetics and physiological chemistry for the years 1946 through 1949.

The coefficients of correlation presented in Table 5 also fluctuate by variable and by year. This variation may be due to random sampling error, actual changes in the makeup of the classes or changes in the curriculum.

Discussion

It would appear that the five tests currently retained in the national testing program of the Council of Dental Education of the American Dental Association are not highly related

to grades earned by students in the University of Minnesota Dental School. The ACE test and the carving test are not predictive of first year honor point ratios in the course areas where they might be expected to have some relationship. Weiss (3) has reported results very similar to those presented here. He found that pre-dental grades and part of the *Survey of Sciences* test gave moderately high correlations with theory and technic grades. Peterson (2) has reported some validity coefficients for this test battery. The correlations which he reports are consistently higher than the ones obtained in the present study. Hence, the present study and that of Weiss illustrate the need for local validation of tests used in national testing programs. Tests which appeared fairly good on a nationwide basis did not show up well in the two dental schools studied.

These studies also show the variability in coefficients of correlation one can obtain when several groups are studied. This means that findings based on one group or a nationwide study should be applied with caution in working with another group for counseling or admission purposes. The use of test data should be tempered by careful consideration of all other available data.

Received August 25, 1952.

References

1. Hansen, L. M. and Paterson, D. G. Scholastic achievement of veterans. *Sch. & Soc.*, 1949, **69**, 195–197.
2. Peterson, S. Forecasting the success of freshman dental students through the aptitude testing program. *J. Amer. Dent. Assn.*, 1948, **37**, 259–265.
3. Weiss, I. Predicting academic success in dental school. *J. appl. Psychol.*, 1952, **36**, 11–14.

THE JOURNAL OF APPLIED PSYCHOLOGY
Vol. 37, No. 4, 1953

Prediction and Practice Tests at the College Level [1]

Scarvia B. Anderson

George Peabody College for Teachers [2]

Attempts to predict college "success" have been so numerous in recent years that it is a rare freshman indeed who has not been subjected to a testing program of one sort or another. The battery of tests used in the present study may be distinguished from hundreds of similar batteries chiefly in that it includes two very easy practice tests which do not seem *prima facie* to belong with the other more conventional placement tests. The questions which we shall attempt to answer here are "What value do these practice tests have in predicting freshman grade point ratio?" and "How does this value compare with the predictive values of the other tests in the battery?"

The original decision to use practice tests in the freshman testing program at George Peabody College for Teachers was based upon the belief that within a group of entering freshmen there would be wide variability in "test-wiseness" and ensuing diminished reliability of test scores. We reasoned, on the basis of observations in previous years, that some of the freshmen would have had no experience with objective tests and that many of the students would not be familiar with the use of special pencils and separate machine-scored answer sheets.[3]

It was felt that a preliminary testing situation would serve two useful purposes: it would aid the students to become more skillful technically, and thus to concentrate their later efforts on the subject matter of the placement tests rather than on the mechanics; and it would perhaps relieve some of their tensions and anxieties.

The tests selected for practice tests were *Otis Quick-Scoring Mental Ability Tests: Beta Tests, Forms Cm and Dm.* These tests were designed for grades 4–9, so it was thought that very few, if any, of the Peabody freshmen would have had any experience with the tests for at least four years.

For each of the forms, Cm and Dm, the time limit stipulated in the *Manual of Directions* is thirty minutes. However, since the students' scores were not to be compared with any norms set up by Otis, time limits of 15 minutes (on the Cm) and 10 minutes (on the Dm) were set. An informal set of instructions appropriate to the groups was substituted for the standardized instructions. These instructions were, of course, identical for groups taking the tests in different rooms. The method of marking the separate answer sheets was explained; the students were told to guess if they wished, but emphasis was placed on the statement that they should not waste time guessing [4]; and the groups were informed that the individual results on the Otis tests would be kept confidential and would not go on their records or to their advisers.

The tests making up the regular freshman battery were given after the practice tests and in this order:

1. *American Council on Education Psychological Examination, 1949 College Edition,* Linguistic Tests (ACE–L); and Quantitative Tests (ACE–Q).

2. *Cooperative English Test C2, Form S,* Vocabulary (C2–V); Speed of Reading Comprehension (C2–S); and Level of Reading Comprehension (C2–L).

3. *Cooperative English Test A:* Mechanics of Expression, Form T (Eng A).

Each individual's total scores and subscores on these three tests, expressed graphically as "normally" spaced percentile ranks based upon local norms, were furnished to the fresh-

[1] Dr. Julian C. Stanley, George Peabody College for Teachers, offered helpful suggestions and criticisms during the preparation of this article.

[2] The author is now associated with the Naval Research Laboratory.

[3] It was determined later from the freshmen who were tested that only one-half had previously used separate answer sheets and special pencils, and that the majority of these had used them only once in a State testing program.

[4] Test scores were corrected for "chance."

Table 1

Intercorrelations and Beta Weights Involved in the Prediction of First Quarter Grade Point Ratio from Freshman Tests
(N = 136)

	0 GPR	1 Cm+Dm	2 ACE-L	3 ACE-Q	4 C2-V	5 C2-S	6 C2-L	7 Eng A
1 Cm+Dm	.48		.75	.62	.59	.67	.57	.67
2 ACE-L	.44			.51	.84	.78	.72	.70
3 ACE-Q	.39				.40	.56	.53	.53
4 C2-V	.25					.69	.65	.60
5 C2-S	.42						.90	.57
6 C2-L	.38							.51
7 Eng A	.54							
Beta Weights		.124	.265	.032	−.427	.104	.075	.413

man advisers as an aid in placement and counseling.

Results

Grade point ratios for one quarter and for three quarters were computed for the students who were included in the study.[5] Excluded from the one-quarter analysis were freshmen who carried less than eight quarter hours and from the three-quarter analysis were freshmen who carried less than 24 quarter hours. A few students, who arrived late or who for other reasons were not present for the first administration of the tests, were given the tests later under less desirable conditions, and it was felt that their test scores were perhaps not comparable with those of the original group. They too were omitted from the study.

The correlations between the seven test scores and between those test scores and GPR for one quarter and for three quarters are shown in Tables 1 and 2.

Note that for the first quarter the correlation between Cm + Dm and GPR exceeds all other correlations with GPR except that of Eng A; for three quarters, r between Cm + Dm and GPR exceeds the correlations of GPR with ACE–Q, C2–V, C2–S, and C2–L. Among these validity coefficients, however, only the differences between Cm + Dm and C2–V for one quarter and between Cm + Dm and ACE–Q for three quarters are significant beyond the one per cent level of confidence,

[5] GPR was computed, letting A + = 12 points per quarter hour, A = 11 points, A − = 10 points, B + = 9 points, and so on down to D − = 1 point, F = 0 points.

Table 2

Intercorrelations and Beta Weights Involved in the Prediction of Three-Quarter Grade Point Ratio from Freshman Tests
(N = 119)

	0 GPR	1 Cm+Dm	2 ACE-L	3 ACE-Q	4 C2-V	5 C2-S	6 C2-L	7 Eng A
1 Cm+Dm	.50		.75	.62	.59	.68	.59	.65
2 ACE-L	.51			.49	.83	.76	.72	.71
3 ACE-Q	.30				.40	.57	.53	.52
4 C2-V	.45					.68	.66	.61
5 C2-S	.43						.90	.57
6 C2-L	.41							.50
7 Eng A	.55							
Beta Weights		.242	.013	−.115	.064	−.082	.166	.367

when the significance test for the difference between r's with a common criterial array is applied.[6]

The regression equations shown below do not meet the requirements of the most economical predictive equations.[7] Rather, the beta weights were computed as a means of studying and comparing the contributions of the various tests to the prediction of GPR. The computation of the beta weights was carried out by a modified Doolittle method.[8]

For one quarter $r_{0\cdot 1234567} = .62$,[9] and the predictive equation, using standard scores, is

$$z_0' = .124z_1 + .265z_2 + .032z_3 - .427z_4 + .104z_5 + .075z_6 + .413z_7.$$

1 refers to Cm + Dm, 2 to ACE–L, 3 to ACE–Q, 4 to C2–V, 5 to C2–S, 6 to C2–L, and 7 to Eng A.

The largest beta weight is that of the Vocabulary part of C2. C2–V seems to serve as a "suppressor" variable. The hypothesis could be advanced that the abilities measured by this test are antithetical to those evaluated by the first-quarter freshman English course teachers at Peabody. Certainly no generality can be implied from this scant evidence, however, especially when one compares this beta weight with the three-quarter beta weight of .064 (see Table 2) and considers "sampling fluctuations."

The various beta weights indicate the relative contributions of the independent variables to the prediction of GPR. Therefore, the contribution of Cm + Dm is almost one-half that of ACE–L, is approximately four times that of ACE–Q, is about 1⅐ times that of C2–S, is approximately 1⅞ times that of C2–L, and is almost one-third that of Eng A.

For three quarters, the multiple R between GPR and the seven test scores is .59. (This R corrected for shrinkage becomes .56.) The predictive equation, using the same notation as above, is

$$z_0' = .242z_1 + .013z_2 - .115z_3 + .064z_4 - .082z_5 + .166z_6 + .367z_7.$$

Here the contribution of Cm + Dm is about two-thirds that of Eng A and exceeds all other contributions.

It is important to mention that "criterion contamination" may be present in the case of all of the predictors except Cm + Dm. The scores on Cm and Dm were the only test scores which were not made available to the teachers, and therefore they could have had no possible influence on the grading. Cm and Dm contributed substantially under fairly ideal conditions.

In arriving at a third predictive equation, the scores on the tests were combined as shown in Table 3.

Table 3

Intercorrelations and Beta Weights Involved in the Prediction of Three-Quarter Grade Point Ratio from Combined Scores on Freshman Tests (N = 119)

	0 GPR	I Cm+ Dm	II ACE Total	III C2 Total	IV Eng A
I Cm+Dm	.50		.80	.68	.65
II ACE Total	.48			.84	.69
III C2 Total	.47				.62
IV Eng A	.55				
Beta Weights		.220	−.090	.171	.363

The multiple R is .59 (.57, corrected for shrinkage). The regression equation becomes

$$z_0' = .220z_I - .090z_{II} + .171z_{III} + .363z_{IV}.$$

I refers to Cm + Dm, II refers to ACE Total, III to C2 Total, and IV to Eng A.

The use of Cm + Dm and Eng A in a fourth predictive equation, using again three-quarter GPR, indicates that the inclusion of Cm + Dm does not greatly enhance the original correlation between Eng A and GPR, though the multiple R, of course, represents a more stable measure. The multiple R is .57. The regression equation becomes

$$z_0' = .212z_I + .412z_{IV}.$$

[6] McNemar (3, pp. 124–125).

[7] See, for example, Thorndike (4, pp. 201 f.).

[8] Thorndike (4, pp. 335–339).

[9] Fisher (1) and Wherry (5) have reported on the amount of bias in a multiple R when it is applied to a new sample and the correlation with a similar criterion computed. In such a situation we may anticipate that the new R will be smaller than the original R. The multiple R of .62, corrected for such shrinkage, is .59. See Kelley's (2) Formula 12:36, p. 474.

where I refers to Cm + Dm and IV refers to Eng A.

It is interesting to note that in every case, Eng A is represented by the highest validity coefficient and the largest beta weight. The fact that English is a required course for freshmen at Peabody and mathematics is not may help explain the larger predictive value of some of the English tests, as compared with ACE–Q; but it does not contribute much toward an explanation of the differences between Eng A and the other English tests. In considering other factors extrinsic to the test itself which might account for some of the high degree of relationship between Eng A and GPR, the fact that Eng A was administered last comes to mind. The time limit (40 minutes) was adequate for far more than half of the testees to finish Eng A; the same was not true of any of the other tests. Perhaps in addition to knowledge of certain English fundamentals, Eng A tested, for this group, motivation and perseverance to a greater degree than any of the other tests. Perhaps, further, these qualities are closely related to freshman GPR at Peabody.

Summary

1. Two easy tests (Otis Quick-Scoring Mental Ability Tests, Beta Tests, for Grades 4–9, Forms Cm and Dm) were administered to a group of entering college freshmen for the purpose of giving the students practice on objective tests and acquainting them with the mechanics of a machine-scored answer sheet.

2. These easy practice tests showed some predictive values above those of several more conventional placement tests; namely, American Council on Education Psychological Examination (1949 College Edition) and Cooperative English Test C2 (Form S). The usefulness of the inclusion of the Otis tests in such a battery is indicated.

3. The Cooperative English Test A: Mechanics of Expression (Form T), in every combination with the other placement tests, contributed most substantially to prediction of freshman GPR.

4. The numerically largest beta weight in the one-quarter regression equation was that of C2–Vocabulary, which in this sample appeared to act as a "suppressor" variable. The corresponding beta weight in the three-quarter predictive equation was small but was positive instead of negative.

Received August 26, 1952.

References

1. Fisher, R. A. Influence of rainfall on the yield of wheat at Rothamstead. *Philos. Trans.*, 1923, **213**, 89–142.
2. Kelley, T. L. *Fundamentals of statistics.* Cambridge: Harvard Univ. Press, 1947.
3. McNemar, Q. *Psychological statistics.* New York: Wiley, 1949.
4. Thorndike, R. L. *Personnel selection: test and measurement techniques.* New York: Wiley, 1949.
5. Wherry, R. J. A new formula for predicting the shrinkage of the coefficient of multiple correlation. *Ann. math. Statist.*, 1939, **2**, 440–457.

THE JOURNAL OF APPLIED PSYCHOLOGY
Vol. 37, No. 4, 1953

Development of a Short Test to Predict a Complex Aggregate Score[1]

Helen Tomlinson and John T. Preston

USAF Training Command, Human Resources Research Center, Personnel Research Laboratory, Lackland Air Force Base, San Antonio, Texas

In a large organization, job assignment involves administration to potential employees of a wide variety of tests to predict probability of success in different kinds of assignments. This is an expensive procedure, and most organizations which require continuous hiring have worked out means of spotting applicants who have little chance to prove successful in any current job openings. One of these is the short preliminary test with cut-off scores based on empirically determined probabilities of meeting qualifying requirements for job appointment.

This article describes a technique for constructing a short test to determine probability of reaching a qualifying score on an aggregate score empirically weighted for prediction of success in one group of training schools. The method was developed in connection with the Service problem of recruitment to meet manpower needs in a specialized job area.

Method

In the Air Force problem, the required short test is to predict probability of reaching a qualifying cut-off on a weighted aggregate score, the aptitude index. Each aptitude index combines the variables of the Airman Classification Battery for optimum prediction of training success in one cluster of related technical specialties.

A preliminary form, twice the length specified for the final predictor test, was composed of blocs of items, with each bloc patterned after one of the components of the aptitude index. Analysis of results from a try-out of the preliminary form determined selection of items for the final form. Test scores included in the aptitude index are converted to normalized standard AF scores before weighting them into the aggregate. Sections of the predictor test are weighted by the number of items so that the total raw score is appropriately weighted. Table 1 shows the composition of the criterion, and the preliminary and final forms of the predictor.

Items in five of the six subtests were selected from appropriate item pools for significant discrimination against the corresponding test of the aptitude index. Experience items were selected and keyed for positive correlation with both the Biographical Inventory score and the aptitude index. Because of the restricted choice of items for Subtest 6, a group of items in a related area, Test VII, was included in the preliminary form for possible use if the items of Subtest 6 did not hold up in the analysis.

The preliminary 64-item form was administered to two unselected samples which had about equal representation of men assigned for basic training to Sheppard and to Lackland Air Force Bases.

For analysis purposes, each sample of 370 was split into the Sheppard group and the Lackland group. Table 2 shows the composition of the samples and score distribution statistics for the preliminary test, the criterion, and a vocabulary test.

Table 1

Composition of the Criterion, the Preliminary Form, and the Final Form of the Predictor

Criterion			Predictor		
				No. of Items	
Test	No. of Items	Weight	Subtest	Prelim.	Final
I	35	2	1	12	6
II	143	2	2	12	6
III	30	1	3	7	3
IV	30	2	4	10	6
V	15	1	5	6	3
VI	30	2	6	10	6
VII	20	–	7	6	–

[1] The views expressed in this article are those of the authors and do not necessarily represent the official views of the United States Air Force.

Table 2

Raw Score Distribution Statistics for the Preliminary Test and Standard AF Score Distribution Statistics for the Aptitude Index and a Vocabulary Test

	N	Preliminary Test M	Preliminary Test SD	Aptitude Index M	Aptitude Index SD	Vocabulary Test M	Vocabulary Test SD	Correlation of Preliminary Test with Aptitude Index
Sample I								
Lackland	194	41.1	11.0	5.5	2.1	6.0	1.8	.86
Sheppard	176	38.2	10.7	5.4	2.1	5.5	1.9	.87
Sample II								
Lackland	193	41.2	10.7	5.6	2.2	5.6	2.0	.86
Sheppard	177	38.3	11.1	5.2	2.1	5.3	2.0	.86

Table 3

Distribution Statistics and Correlations of the Predictor* with the Criterion

	N	Predictor M	Predictor SD	Aptitude Index M	Aptitude Index SD	Correlation of Predictor with Aptitude Index
Sample I						
Lackland	194	19.5	5.8	5.5	2.1	.87
Sheppard	176	17.7	5.9	5.4	2.1	.85
Total	370	18.7	5.9	5.5	2.1	.85
Sample II						
Lackland	193	19.4	6.1	5.6	2.2	.86
Sheppard	177	17.8	6.0	5.2	2.1	.84
Total	370	18.6	6.1	5.4	2.1	.85
Sample III (Independent)	882	17.0	6.6	4.9	2.1	.86

* Scoring keys for the selected 30 items applied to answer sheets for the preliminary test.

Selection of items for the final form was based on several criteria: (1) Correlation (Ø) with the aptitude index; (2) Correlation (Ø) with the parent test; (3) Lower discrimination for other tests of the aptitude index composite than for the parent test; and (4) Difficulty index centering around 60% R (uncorrected).

Effect of chance error was diminished by running the item statistics separately for the two samples. Instead of preparing separate 30-item keys for each sample, the criterion of consistency between samples was added for producing a single 30-item key.

Efficiency of Prediction

Application of the 30-item key to the experimental answer sheets yielded the distribution statistics and correlations appearing in Table 3. The key was developed on Samples I and II. Sample III is independent.

Since the new test is essentially a short form of the aptitude index, the predictor has a special relation to the criterion. The Kuder-Richardson estimate of reliability (Formula

Table 4

Correlations with Control Aptitude Index (Sample: 324 Lackland Airmen)

	M	SD	Correlation with Control Aptitude Index
Control Aptitude Index	5.8	2.1	
Criterion Aptitude Index	5.7	2.1	.77
Predictor Score	19.9	5.7	.65

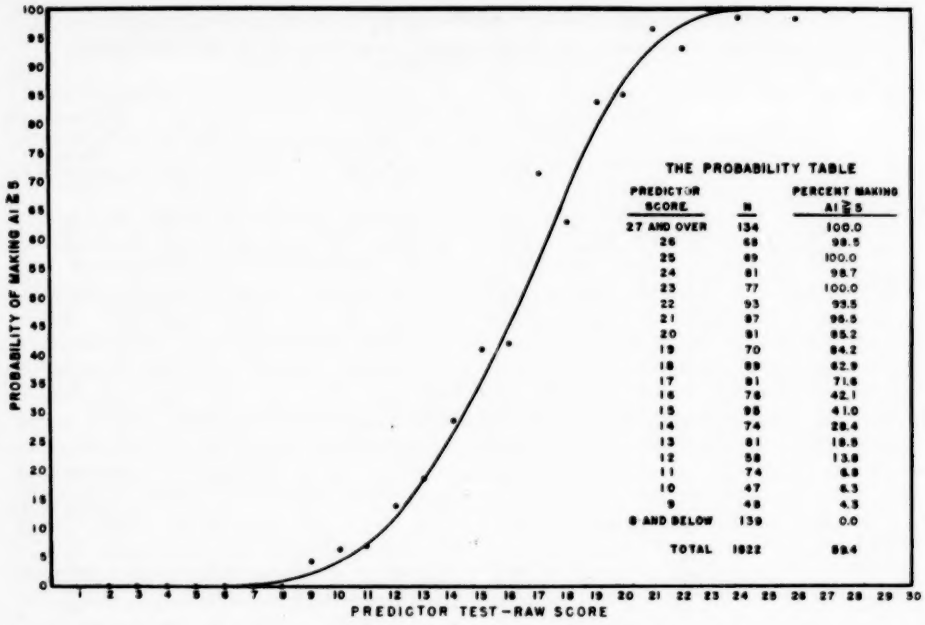


FIG. 1. Curve showing probability of making an aptitude index of 5 or more for each raw score on the predictor test.

20) is .84. The estimate obtained by correlating scores for the final form against scores for the remaining 34 items is .82. These reliability coefficients do not differ significantly from the coefficients of prediction (.84 — .87).

To show that the predictor test is specifically predictive of the criterion, the correlation was computed with a control aptitude index of median correlation with the criterion aptitude index and maximum correlation with the Armed Services Qualification Test. Table 4 shows these correlations.

Because of common components in the eight Air Force aptitude indexes, their intercorrelations are high. Consequently the correlation of the predictor with the control aptitude index (.65) is less than the correlation between the criterion aptitude index and the control aptitude index (.77).

Rather than a direct score conversion table, empirical probabilities for achieving an aptitude index of at least 5 were computed from the combined sample of 1622 for each predictor score directly from the 30 x 9 regression chart. Figure 1 shows that the curve of percentages closely approximates the expected ogive. The steepness of the curve between predictor scores 16 and 17 suggests a reliable cut-off for a better-than-even chance of qualifying for training in the specific job area.

Summary

This method of constructing short screening tests offers two advantages:

1. The screening process can be readily directed to current personnel requirements.

2. A simple "Rights" score read into a table of probabilities makes it possible to use clerical assistants with only a minimum of training for administering, scoring, and interpreting the tests.

The method is applicable only in conjunction with reliable, well-validated selection instruments.

Received November 10, 1952.

THE JOURNAL OF APPLIED PSYCHOLOGY
Vol. 37, No. 4, 1953

The Classification of Occupations by Means of Kuder Interest Profiles: I. The Development of Interest Groups

John L. Holland, Allen H. Krause, M. Eloise Nixon, and Mary F. Trembath

Vocational Counseling Center, Western Reserve University

The need for a more extensive knowledge of occupational interests is well recognized by vocational counselors. The empirical evidence is limited largely to the data concerning the Strong Vocational Interest Blank and the Kuder Preference Record. Since these inventories have been applied to only a small number of occupations, it is frequently necessary in practice to speculate about the nature of the interests for many occupations. An "unknown" occupation presents at least two problems: (1) What are the characteristic interests of this occupation?, and (2) In what known interest group does this occupation belong?

These problems are commonly approached by developing interest keys using Strong's method or by securing Strong interest profiles for unknown occupations. Since these methods present practical and financial problems, a simpler, less expensive method of measuring and classifying interests for a large number of occupations would accelerate the extension of our knowledge of interests and interest groups.

This study is an attempt to classify occupations empirically using KPR profiles and to present a research tool for validation. While the sets of interest groups developed here are empirical in nature, their predictive power, especially with respect to job satisfaction, is largely unknown. It is hoped that others will be able to test and extend this classification system by making cross-comparisons with the Strong, and by predicting job satisfaction or vocational choice.

Method

A group of 45 KPR profiles representing the major interest groups for men, as defined by the SVIB, was selected from the Kuder manual of instructions (4). This selection was made since the Strong groups represent a classification system supported by considerable evidence including a number of factor analyses by Strong (6) and Thurstone (8). It is assumed that a new method of classification would reveal comparable interest groups.

About 27 of the 45 occupations used in this sample are similar to, if not identical with, the Strong sample of 44 occupations. Eighteen additional Kuder profiles were used to define more accurately the limits of a given interest group, and to increase the representativeness of the group by employing a larger sample of occupations. This need is especially apparent in the case of groups defined by one to three occupations.

A second group of 42 profiles was secured for women. No attempt was made to select these profiles in terms of occupational groups; however, apprentice and military occupations, except for aviation assembly and repair worker, were avoided. Also, housewives as an occupational group were not used since the Kuder sample is limited to only three groups of housewives (wives of lawyers, physicians, and farmers) and may accordingly be quite unrepresentative of housewives-in-general.

For females, the Kuder sample contains about 17 of the 25 occupations in the present Strong Blank. The twenty-five additional occupations again were used to define the limits and to increase the representativeness of the interest groups.

Using *rho* as an index of profile similarity, the Kuder profiles were intercorrelated.[1] By grouping profiles which intercorrelate .70 or greater, a set of interest groups was derived

[1] Profiles were coded by listing the highest scale first with the remaining scales following in descending order. For these computations the outdoor scale was omitted since it was available for only a few occupations. The matrices formed by the intercorrelation of profiles (16 pages) are deposited with the American Documentation Institute. Order Document 3725 from American Documentation Institute, Library of Congress, Washington 25, D. C., remitting $1.00 for microfilm (images 1 inch high on standard 35 mm. motion picture film) or $2.40 for photocopies (6 × 8 inches) readable without optical aid.

for each sex. This procedure consisted of a simple cluster analysis. Each profile was inspected for its pattern of intercorrelations. Occupational groups were formed by classifying profiles having similar intercorrelational patterns. In most cases the evidence for a particular grouping is clear-cut; however, when a profile correlates with about equal frequency with the occupations of two or more groups, it is listed in each group in order to give an accurate picture of the data.

It seemed meaningful to make several exceptions to this method of classification. In the case of males, the profiles for farmer, aviator, carpenter, and forest supervisor are placed in the interest group with which they show the highest single correlation. For females, this relaxation of the criterion was made for stenographer and typist, and sales clerk. Although these profiles fail to meet the criterion for inclusion, their intercorrelational patterns are similar to the typical pattern of their respective groups.

In order to secure a useful arrangement of occupations within a given interest group, occupations are ordered in terms of their "representativeness." The occupation which shows the greatest number of correlations equal to or greater than .70 with the other members of the same group is listed first in that group. This occupation is designated as the "core" occupation. The remaining occupations are then arranged in descending order of correlation with the core profile. Ties in rank were resolved by comparing the number of correlations equal to or greater than .70 for each of these occupations with the other occupations in the same group. The relative position of

Table 1

Kuder Interest Groups (Male)

Group	Occupation	Rho	Code*	N
I.	Skilled and Technical			
	Asst. District Rangers		0'1325687'94	28
	Electricians	.95	1'30528697'4	37
	Vocational Training Teachers	.87	'15382697'4	35
	Machinists	.80	'153279864'	117
	Meteorologists** (2)	.80	3'6127589'4	185
	Engineers (3)	.78	'3216547098'	653
	Draftsmen	.77	5'13279684'	216
	Aviators	.52	31'796528'4	34
	Farmers	.39	0'913825647'	129
	Carpenters	.32	'1058973246'	124
II.	Managerial			
	Production Managers		'2631495807'	139
	Engineers (3)	.79	'3216547098'	653
	Controllers	.75	23'916475'8	24
III.	Scientific			
	Psychologists		36'285791'4	111
	H. S. Teachers of Mathematics	.87	23'867519'4	30
	Laboratory Technicians	.81	3'56278914'	90
	All Physicians and Surgeons	.78	3'658071294'	260
	Chemists	.76	3'672518'49	54
	District Rangers	.73	0'256381749'	102
	Meteorologists (2)	.68	3'6127589'4	185
	Protective Service Occupations	.66	'6870351249'	23
	Engineers (3)	.43	'3216547098'	653
IV.	Drugstore Managers and Pharmacists		'342987561'	140

* Scales to the left of the first apostrophe indicate scales at or above the 75th percentile. Scales to the right of the last apostrophe are at the 25th percentile or below.

** The number in parentheses following occupational title designates occupations listed in two or three interest groups.

Table 1—*Continued*

Group	Occupation	Rho	Code*	N
V.	Welfare			
	H. S. Teachers of Social Studies		8'6243795'1	23
	School Administrators	.88	8'62934751'	65
	Social and Welfare Workers	.83	8'6439257'1	53
	Forest Supervisors	.68	0'682543197'	17
	Clergymen	.63	86'793254'1	43
VI.	Clerical			
	General Office Clerks		9'267438501'	110
	General Accountants	.98	29'64730851'	92
	Office Mgrs. and Chief Clerks	.91	2'964758301'	138
	Cost Accountants	.91	29'6475318'	28
	Printers and Pressmen	.83	'976253418'	32
	Financial Institution Clerks	.78	2'98674351'	25
	Lawyers and Judges (2)	.75	6'97428503'1	331
	Purchasing Agents and Buyers (2)	.73	'469278531'	103
VII.	Expressive			
	Sales Managers		4'768953201'	230
	Life Insurance Salesmen	.98	4'7869532'1	50
	Salesmen, to Consumers	.96	4'678935012'	353
	Reg. Salespersons, Dept. Store	.90	4'7965823'1	184
	Stock and Bond Salesmen	.90	4'8796523'1	118
	Advertising Agents	.88	64'75893'12	26
	Route Salesmen	.87	4'87569312'	104
	Authors, Editors, and Reporters	.85	6'794850'321	113
	Personnel Managers	.83	'486792305'1	50
	Banking, Finance, and Insurance Officials	.82	'4796258031'	42
	Lawyers and Judges (2)	.75	6'97428503'1	331
	Purchasing Agents and Buyers (2)	.70	'469278531'	103
	Musicians and Music Teachers	.70	7'65894'321	77
	Commercial Artists	.50	5'6794281'3	31

an occupation within its group serves then as a crude index of its communality or belongingness.

Each set of interest groups was arranged by inspecting the matrix formed by the core occupations. The highest negative correlation between a pair of core occupations designates the first and last interest groups, or the groups whose interests are most divergent. The remaining groups are placed between these groups in accordance with their correlation with the first group listed.[2]

Results

Males. The classification of profiles produced seven interest groups which have been designated tentatively as: I. Skilled and Technical; II. Managerial; III. Scientific; IV. Drugstore Managers and Pharmacists; V. Welfare; VI. Clerical; and VII. Expressive. Table 1 shows the male sample classified into these groups. To increase the usefulness of Table 1, the *rho* with the core occupation, the N, and the code [3] for each occupation have been listed.

In general, the obtained groups are similar to related Strong Groups; the Kuder Groups I, II, III, V, and VI agree in content with the Strong Groups IV, III, I, V, and VIII, respectively. The differences between comparable groups seem slight.

[2] If the remaining groups are arranged in accordance with their correlation with the last group, groups V and VI will change position, but the original arrangement is essentially the same.

[3] The attention of readers is directed to the misleading character of a mechanical coding device *for counseling purposes* that isolates percentiles of 75 or greater or percentiles of 25 or lower. See Diamond, S. The interpretation of interest profiles. *J. appl. Psychol.*, 1948, **32**, 512–520.—*Editor.*

Table 2

Kuder Interest Groups (Female)

Group	Occupation	Rho	Code*	N
I.	Computational			
	H. S. Teachers of Mathematics		2'31597864'	47
	Office Machine Operators	.93	2'13598746'	62
	Tearoom and Restaurant Mgrs.	.73	2'3514697'8	20
	Hospital Dietitians** (2)	.70	3'21765849'	31
	Statistical Clerks	.43	2'36574189'	27
II.	Scientific and Technical			
	H. S. Teachers of Home Economics		'513867249'	136
	Aviation Assembly & Repair Wrkrs.	.90	1'3578624'9	75
	Cooks and Bakers	.88	'16347582'9	31
	Physicians	.87	31'85627'49	43
	All Trained Nurses	.83	8'35176249'	1071
	Supervisors and Head Nurses	.70	'831752649'	196
	Dental Hygienists	.70	83'571462'9	35
	Laboratory Technicians	.70	31'527486'9	31
	Occupational Therapists	.63	15'738624'9	70
	Home Demonstration Agents	.62	'85147326'9	24
	Hospital Dietitians (2)	.43	3'21765849'	31
III.	Clerical			
	H. S. Teachers of Commercial Subjects		2'94756138'	64
	General Office Clerks	.85	'295471368'	136
	Bookkeepers	.80	2'96754381'	62
	Stenographers and Typists	.32	'973562148'	235
IV.	Linguistic			
	Journalists		6'5473182'9	31
	Copy Writers, Mail Order Co.	.95	64'57132'89	19
	Librarians	.78	6'57149328'	39
	Artists and Art Teachers	.75	5'167483'29	22
	H. S. Teachers of Language (2)	.75	6'75489132'	42
	H. S. Teachers of English (2)	.71	6'75482193'	110
V.	Expressive			
	Assistant Buyers, Dept. Store		46'7852193'	58
	Personnel Mgrs., Mail Order Co.	.87	4'68751932'	34
	H. S. Teachers of English (2)	.86	6'75482193'	110
	H. S. Teachers of Social Studies	.80	'684527193'	56
	Social Workers	.77	8'647531'29	50
	H. S. Teachers of Language (2)	.75	6'75489132'	42
	Personnel Wrkrs. other than Mgrs.	.73	48'6217953'	27
	Musicians and Music Teachers	.70	7'5684192'3	68
	Cashiers	.70	74'6529381'	43
	Retail Buyers	.70	4'7625139'8	29
	Salespersons, Dept. Store, Reg.	.70	4'75896231'	617
	Religious Workers	.68	8'6754139'2	31
	Floor and Section Mgrs., Dept. St.	.65	4'68297135'	25
	Secretaries	.65	'467952138'	121
	Executives, Mail Order Company	.56	4'27653981'	67
	Teachers, Primary and Kindergarten	.50	'857621493'	544
	Office Mgrs., Chief Clerks	.40	'426179358'	29
	Sales Clerks	.38	'849761325'	26
	Telephone Operators	.22	'495781236'	22

* Scales to the left of the first apostrophe indicate scales at or above the 75th percentile. Scales to the right of the last apostrophe are at the 25th percentile or below.

** The number in parentheses following occupational title designates occupations listed in two or three interest groups.

In the Kuder classification, engineers are classified in the skilled, scientific, and managerial groups rather than in science alone as in the SVIB. District rangers are classified in both the skilled and scientific groups rather than in science alone. Protective service occupations are classified in science rather than in skilled trades. Personnel managers are classified in the expressive group rather than social service. Lawyers are classified in the clerical and expressive groups rather than language occupations only. Printers and pressmen are classified in the clerical group rather than in the skilled trades. Purchasing agents and buyers are classified in both the clerical and expressive groups rather than in business detail alone. Pharmacists and drugstore managers form a separate group in the Kuder data while they fall in Group VIII on the Strong.

These differences are due in part to the classifying of occupations in one or more groups by the KPR, and to the limiting of an occupation to a single group by the SVIB. Consequently, only four of the above occupations are placed in opposed interest groups: protective service occupations, personnel managers, printers and pressmen, and pharmacists and drugstore managers.

The creation of Group VII, Expressive Occupations, presents the greatest difference between these interest systems. Group VII incorporates Strong Groups IX and X and a number of other occupations: salesmen to consumers; regular salespersons (department store); stock and bond salesmen; route salesmen; personnel managers; banking, finance, insurance officials; purchasing agents and buyers; musicians and music teachers; and commercial artists. The cohesiveness of Group VII is marked. The first four occupations intercorrelate .70 or greater with each other. Furthermore, the median *rho* for the entire group of 14 occupations with the core occupation, sales manager, is .86.

While it is possible to fractionate Group VII into three sub-groups consisting of sales, language, and art (including music), there is no empirical justification for such groups since the correlational patterns show no sharp differences. This procedure would, of course, make the two interest systems appear more alike.

Females. The correlational matrix for females produced five interest groups: I. Computational; II. Scientific and Technical; III. Clerical; IV. Linguistic; and V. Expressive. The comparability of these groups with the 13 groups shown by Strong (7) is difficult to ascertain. The Strong groups are defined frequently by a single occupation so that their limits are not clearly structured.

Kuder Group I, Computational, includes Strong Group XII (mathematics and science teacher) and three additional occupations not found in the Strong sample: office machine operators, tearoom and restaurant managers, and statistical clerks. Hospital dietitian in Strong Group IX falls in this group and also in Group II, Scientific and Technical.

Kuder Group II, Scientific and Technical, includes Strong Groups IX, XI, and XIII (home economics teacher, dietitian, occupational therapist, nurse, laboratory technician, and physician) as well as a number of occupations not contained in the SVIB.

Kuder Group III, Clerical, is a rough approximation of Strong Group VII as both groups include high school teachers of commercial subjects, general office workers, bookkeepers, and stenographers.

Kuder Group IV, Linguistic, combines Strong Groups I and II (artist, librarian, and English teacher) and a number of language occupations not included in the SVIB.

Kuder Group V, Expressive, combines Strong Groups III, IV, VI and VIII (social worker, social science teacher, elementary school teacher, and buyer) as well as a number of related sales, social service, and administrative occupations. High school teachers of English which occur in Kuder Group IV are also listed in Group V.

In general, the Kuder set of interest groups appears to be a coarser classification system than the Strong set, but the direction of their classification appears essentially the same. A more accurate picture of the relationships between these systems can be obtained by a comparison of their correlational matrices.

Discussion

The occupational classification system presented here is limited by the original data. The representativeness of most of the occupational profiles is largely unknown; furthermore, each criterion group contains both successful and unsuccessful, experienced and inexperienced, as well as satisfied and dissatisfied workers (4). The usefulness of this interest system in interpreting individual profiles is consequently restricted.

The validity of these interest groups is problematical. Tests of significance are inappropriate because of the large number of computations and because of the unorthodox use of *rho;* that is, varying N's and yet only eight degrees of freedom for each correlation. The use of *rho* as an index of profile similarity violates the assumptions underlying its application. Moreover, as Guilford (3) points out, the "ipsative property of the Kuder scores . . . renders their use for intercorrelations among themselves . . . so questionable as to preclude attempts at analysis by the R-technique."

These difficulties pose a choice of abandoning the evidence until more adequate data can be obtained and statistical elegance can be achieved, or to test the interest groups by means of a prediction study. The latter alternative seems desirable in view of the need for an immediate understanding of the present data. Although a predictive test may make statisticians wince, it will yield an estimate of the predictive power of the classification.

While they furnish only suggestive and incomplete evidence, several related studies support the Kuder interest groups. In a factor analysis of the SVIB, KPR, Bell Inventory, and MMPI, Cottle (2) extracted seven factors similar to those described by Thurstone (8) and Strong (6). Cottle's interest factors support the Kuder classification including Group VII, Expressive, which combines Strong Groups IX and X in addition to a number of other occupations. The factor G obtained by Cottle appears similar to the Kuder Group VII since it is characterized by high positive loadings on Strong Groups IX and X and high negative loadings on the Kuder mechanical scale. Similarly, an inspection of the present Kuder matrix indicates that the expressive occupations correlate most negatively with the skilled trade occupations.

In addition, Cottle's matrix for the SVIB with the KPR reveals that the Strong Group keys correlate with the KPR scales in a pattern which is similar to the coded profile for the comparable Kuder interest group. The positive and negative correlations of Kuder scales with a given Strong group scale produce coded profiles which are similar to those in the Kuder Groups. For example, the code for high school teachers of social studies, the core occupation for Kuder Group V, Welfare, is 862437951. Cottle's matrix produces the code 867495213. The other five group scales show similar relationships with the Kuder interest groups.

In a study of occupational test patterns, Barnette (1) has supplied Kuder profiles for "success" and "failure" samples for five occupational groups: engineers, accountants, clerical personnel (except accountants), a specialized clerical group (primarily verbal in nature), and salesmen. When the pairs of "success" and "failure" profiles are correlated with comparable Kuder core occupations, six of the seven tests reveal that the "success" groups correlate higher with the core occupation than do the "failure" groups. The data

Table 3

The Relation of Occupational Success and Failure to Core Occupations of Corresponding Kuder Interest Groups*

Occupations		N	I	II	III	VI	VII
Engineers	S	83	.83	.66	.44		
	F	39	.59	.21	−.09		
Accountants	S	74				.92	
	F	24				.85	
Clerical (Specialized)	S	20				.78	
	F	14				.47	
Clerical (General)	S	54				.83	
	F	40				.48	
Salesmen	S	77					.86
	F	56					.94

* Correlations are *rho* coefficients.

suggest that the degree of similarity of an individual profile to a given interest group is related to success or failure in the occupation. Table 3 shows these relationships.

The classification of occupations by means of the entire profile appears to have several advantages which other systems such as the method proposed by Weiner (12) lack. Systems of classification which rely on the two or three highest scales, or the scales above a certain cutting score, utilize only a portion of the data. There is no experimental evidence revealing that "high" scores are more significant or useful than "low" scores. In practice, a coding system which uses both low and high scores impels the counselor to notice and use more of the data. Finally, a total code minimizes the misclassification of occupations. If only the high scores are used to categorize an occupation, occupations which have similar patterns may be sorted into different groups by means of one or two scales which may represent the only differences between a pair of profiles.

The results suggest further that previous studies of the SVIB and KPR (**5, 9, 10, 11, 13**) may be in error in concluding that these instruments show little relationship. Their findings may be largely negative due to an inappropriate level of analysis; that is, a single scale comparison rather than a pattern method.

Summary

A sample of KPR profiles was intercorrelated and classified by means of a simple cluster analysis. Sets of occupational groups for men and women were derived which are comparable with many of the SVIB groups. While these groups are empirical in nature, their validity is unknown. It was suggested that validation might be established by prediction studies of job satisfaction and vocational choice.

Received August 29, 1952.

References

1. Barnette, W. L. Occupational aptitude patterns of selected groups of veterans. *Psychol. Monogr.*, 1951, **65**, No. 5 (Whole No. 322).
2. Cottle, W. C. A factorial study of the Multiphasic, Strong, Kuder, and Bell inventories using a population of adult males. *Psychometrika*, 1950, **15**, 25–47.
3. Guilford, J. P. When not to factor analyze. *Psychol. Bull.*, 1952, **49**, 26–37.
4. Kuder, G. F. *Examiner manual for the Kuder Preference Record, vocational, form C, second revision.* Chicago: Science Research Associates, March 1951.
5. Peters, E. F. Vocational interests as measured by the Strong and Kuder inventories. *Sch. and Soc.*, 1942, **55**, 453–455.
6. Strong, E. K., Jr. *Vocational interests of men and women.* Stanford: Stanford Univ. Press, 1943.
7. Strong, E. K., Jr. *Manual for vocational interest blank for women.* Stanford: Stanford Univ. Press, 1951.
8. Thurstone, L. L. A multiple factor study of vocational interests. *Personnel J.*, 1931, **10**, 198–205.
9. Triggs, F. O. A study of the relation of Kuder Preference Record scores to various other measures. *Educ. Psychol. Measmt.*, 1943, **3**, 341–354.
10. Triggs, F. O. A further comparison of interest measurement by the Kuder Preference Record and the Strong Vocational Interest Blank for men. *J. educ. Res.*, 1944, **37**, 538–544.
11. Triggs, F. O. A further comparison of interest measurement by the Kuder Preference Record and the Strong Vocational Interest Blank for women. *J. educ. Res.*, 1944, **38**, 193–200.
12. Wiener, D. N. Empirical occupational groupings of Kuder Preference Record profiles. *Educ. Psychol. Measmt.*, 1951, **11**, 273–279.
13. Wittenborn, J. R., Triggs, F. O. and Feder, D. D. A comparison of interest measurement by the Kuder Preference Record and the Strong Vocational Interest Blanks for men and women. *Educ. Psychol. Measmt.*, 1943, **3**, 239–257.

THE JOURNAL OF APPLIED PSYCHOLOGY
Vol. 37, No. 4, 1953

The Validity of the Mooney Problem Check List *

Charles J. McIntyre

The Pennsylvania State College

The Problem Check List is an instrument developed by Mooney (1) to enable the teacher or counselor to quickly identify problems or problem areas which concern his students. The high school form, which was used in this study, consists of 330 problems found to be of particular concern to students. They are classified into the following eleven major areas: (1) Health and Physical Development; (2) Finances, Living Conditions, and Employment; (3) Social and Recreational Activities; (4) Courtship, Sex, and Marriage; (5) Social-Psychological Relations; (6) Personal-Psychological Relations; (7) Morals and Religion; (8) Home and Family; (9) The Future: Vocational and Educational; (10) Adjustment to School Work; and (11) Curriculum and Teaching Procedures.

Normally the subject is instructed to underline those problems that bother him and to circle those underlined problems which trouble him the most. In this study no distinction was made between underlined and circled items.

Mooney (2) has said that the nature of the Check List makes it impossible to arrive at a definitive conclusion about its validity. Validity, he says, must be determined in terms of the particular purpose and the particular situation. While it probably is true that conventional measures of validity are difficult if not impossible to obtain for an instrument of this kind, it appears nevertheless that the Check List should meet at least three minimum requirements: (1) Students recognize their own problems; (2) They find these problems listed on the Check List; and (3) They are willing to record them.

This study assumes that if these three conditions are met it should be possible to predict the relative number of problems listed by particular groups of students in particular areas. Hence the following hypotheses were formulated: (1) The less intelligent students would have more problems than the more intelligent in the area of Adjustment to School Work; (2) Seniors would have more problems than those in the lower grades in the area of The Future: Vocational and Educational; (3) Students from broken homes would have more problems than those from intact homes in the area of Home and Family; (4) Boys would have more problems than girls in the area of Adjustment to School Work; (5) Boys would have more problems than girls in the area of The Future: Vocational and Educational; (6) Negroes would have more problems than whites in the area of Finances, Living Conditions, and Employment; and (7) Girls would have more problems than boys in the area of Courtship, Sex, and Marriage.

The rationale behind each of the hypotheses should be evident.

Procedure

Subjects. The subjects were 407 high school students in grades ten to twelve inclusive. The school which they attended was the only public high school in a highly industrial Pennsylvania city with a population of approximately sixty thousand. The city population is highly heterogeneous in terms of race, religion and national origins, and this heterogeneity is reflected in the school population.

Method. Approximately one-fourth of the school population was sampled. The Check Lists were administered by the homeroom teachers during the period prior to the first class in the morning. Homerooms were selected so that the proportion of students in the several courses and classes in the sample would approximate the proportion of students in these courses and classes in the entire school population. In this way it was possible to secure a reasonably representative sample

* This paper represents the substance of a thesis submitted in partial fulfillment of the requirements for the M.S. degree at The Pennsylvania State College.

Table 1

Groups and Problem Areas Relevant to Hypotheses 1 to 6 with N, Mean, SD and CR for Each

Hypothesis	Problem Area	Group	N	Mean	SD	CR
1.	Adjustment to School Work	Less intelligent	55	5.1	.28	3.40
		More intelligent	61	2.9	.12	
2.	The Future:	Seniors	156	3.5	.07	3.39
	Vocational and Educational	Sophomores	157	2.4	.04	
3.	Home and Family	Broken Home	85	2.8	.10	2.49
		Intact Home	318	1.9	.02	
4.	Adjustment to School Work	Boys	202	4.6	.07	2.28
		Girls	204	3.8	.05	
5.	The Future:	Boys	202	3.3	.05	2.14
	Vocational and Educational	Girls	204	2.6	.04	
6.	Finances, Living Conditions	Negro	100	3.5	.08	2.11
	and Employment	White	295	2.8	.03	

while retaining the administrative convenience of intact homerooms.

In questionnaires of this kind the problem of a student's honesty is a serious one, particularly when there is a chance, as here, that his teachers may check his responses. As an example Olson (3), using the Woodworth-Mathews Personal Data Sheet, found that more symptoms were reported when the questionnaire was left unsigned. Therefore in this study a supplementary instruction sheet was attached to the Check List explaining that the study was being conducted to gather information on the problems of high school students and instructing them not to sign the Check List. They were, however, to put their names on the instruction sheet and hand this in for an attendance record. A system of discrete pinholes pricked through both the instruction sheet and the Check List made it possible to later match the two and identify the Check List.

Information on the relevant variables was abstracted from the students' records in the school file. Only two of these variables require further definition: (1) "Less intelligent students" are defined in this study as those students whose Otis Gamma IQ's were more than one standard deviation below the mean IQ of the sample. "More intelligent students" are those whose IQ's were more than one standard deviation above the mean of the sample; and (2) A student was classified as coming from a broken home if the records indicated that he was not presently living with both natural parents.

Treatment of the Data. The mean number of problems reported was computed for each of the variables and problem areas which was relevant with respect to the hypotheses. The hypotheses were tested by computing the critical ratio of the difference between these means.

Results

1. Hypotheses 1 and 2 were confirmed. In Table 1 it will be seen that the differences between the mean number of problems reported in each case were significant at or beyond the .01 level of confidence.

2. Hypotheses 3, 4, 5, and 6 were confirmed. In Table 1 it will be seen that the differences between means were significant at the .05 level of confidence.

3. Hypothesis 7 was not confirmed. No statistical difference between means was found.

Summary and Conclusions

The problem of determining the validity of the high school form of the Mooney Problem Check List was attacked by computing the mean number of problems checked in particular problem areas by a group of high

school students who were classifiable into various discrete groups.

This study was founded upon the assumption that the essential test of the validity of an instrument of this kind consists in determining whether or not the students can recognize their own problems, find these problems represented on the Check List, and record them. If these three criteria are met the mean number of problems checked in particular areas by various groups should differ significantly in a reasonable and predictable way.

Hence, seven such differences were hypothesized on rational grounds. That is, because of the sociological and psychological characteristics of particular groups, it was predicted that some groups would check more problems in certain areas than other groups, providing the three criteria of validity specified above were met by the Check List. Of the seven differences hypothesized, six were found.

It is concluded that these findings present prima facie evidence for the validity of the Check List.

Received September 2, 1952.

References

1. Mooney, R. L. Exploratory research on students' problems. *J. educ. Res.*, 1943, **37**, 218–224.
2. Mooney, R. L. and Price, Mary. *Manual to Accompany Mooney's Problem Check List—High School Form*. Columbus, Ohio: Ohio State University, 1948.
3. Olson, W. C. The waiver of signatures in personal data reports. *J. appl. Psychol.*, 1936, **20**, 442–450.

THE JOURNAL OF APPLIED PSYCHOLOGY
Vol. 37, No. 4, 1953

A Comparison of Manual and College Norms for the MMPI

Fred. T. Tyler and John U. Michaelis

School of Education, University of California, Berkeley, California

Various investigators have suggested that the distributions of scores of college students on certain of the MMPI scales are different from those reported in the manual of directions. For instance, McKinley (2) reported that college students obtained higher K-scores than did a less selected (intellectually and educationally) population. Similarly, college men have been found to score high on the masculinity-femininity scale (3), indicating a "deviation of the basic interest pattern in the direction of the opposite sex" (1, p. 5). The MMPI has been extensively used in research and clinical investigations at the college level, so that it should be of interest to know something of the extent to which T-scores obtained from the standardization group correspond to those based upon a college population.

The MMPI (Booklet—short form) was administered to nearly one thousand juniors, seniors and first-year graduate students in education courses in the School of Education at the University of California, either in Education 110 (Educational Psychology) or in those courses required in the programs leading to the General Elementary and General Secondary Teaching Credentials. T-scores were computed for each of the nine clinical scales [1] for men (N = 470) and women (N = 571) separately. These samples are considerably larger (about 60 and 40 per cent respectively) than those of the original standardization groups. It is recognized that these norms might not be representative of the general college population, since the majority of the subjects were expecting to enter the teaching profession. However, the college norms to be discussed here may be of interest to those who are administering this Inventory to college students, and especially to those concerned with personnel problems in a teacher education program.

It is the purpose of this note to present a brief comparison between the norms reported by Hathaway and McKinley (1) and by Tyler and Michaelis (6). Comparative data are presented in Table 1 for women and Table 2 for men.

For the women, there seem to be some differences between the norms on three scales,

[1] The scores were not corrected by means of the K-score (2, 7).

Table 1

Raw Score Equivalents of Selected Standard Score Values Based on Manual and U. C. Norms for Female Subjects

	Standard Scores											
	80		70		60		50		40		30	
Scales	M*	C†	M	C	M	C	M	C	M	C	M	C
Hs	23	14	17	11	12	7	7	4	1	1		
D	35	33	29	29	24	24	19	19	14	14	9	10
Hy	36	34	30	30	24	26	19	21	13	17		
Pd	26	26	22	22	18	18	13	14	9	10	5	6
Mf	22	24	27	29	31	33	36	37	41	42	46	46
Pa	18	17	15	14	11	12	8	9	5	6	1	4
Pt	36	30	28	24	21	17	13	10	5	4		
Sc	35	28	27	21	19	15	11	9	3	3		
Ma	27	27	23	23	19	19	14	15	9	12	5	8

* Raw score equivalents based on norms in the Manual (1).
† Raw score equivalents based on college norms (6).

Table 2

Raw Score Equivalents of Selected Standard Score Values Based on Manual and U. C. Norms for Male Subjects

	Standard Scores											
	80		70		60		50		40		30	
Scales	M*	C*	M	C	M	C	M	C	M	C	M	C
Hs	18	14	13	10	9	7	4	4	0	0		
D	29	34	25	28	21	23	17	18	12	13		
Hy	33	32	27	28	22	24	16	20	11	16		
Pd	26	27	22	23	18	19	14	15	10	10	6	6
Mf	36	45	30	39	25	33	20	27	15	21	10	15
Pa	18	17	15	14	11	12	8	9	5	6	1	4
Pt	31	30	24	23	17	16	10	9	2	1		
Sc	32	30	24	23	17	15	9	8	2	1		
Ma	27	28	23	24	18	20	14	16	9	12	5	8

* See footnote to Table 1.

Hs, Pt and Sc. Using the manual norms as the basis for interpretation, it is suggested that the college women were less concerned about their health, were freer from fears and lack of confidence, and from bizarre or unusual thoughts (1, 4). On the remainder of the scales there was a high degree of similarity between the norms obtained from the two samples.

Several comments about the two sets of norms are in order:

1. The college men obtained relatively lower scores than did the standardization sample on only one scale, Hs. Accepting the usual interpretations of this scale, it appears that the college men were less concerned than were these members of the standardization group about the state of their health.

2. The college men appeared to be somewhat less depressed and more feminine in their interests when compared with the members of the original standardization group. For instance, on the D scale, a raw score of 29 has a T-score equivalent of 80 on the manual norms and of 71 on the college norms. Similarly on the Mf scale, a raw score of 30

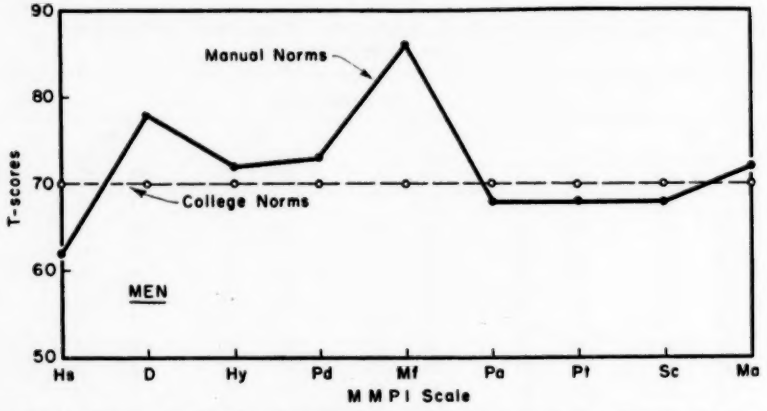


FIG. 1. T-scores on the Manual norms of raw scores corresponding to T-scores of 70 on the College norms (Men).

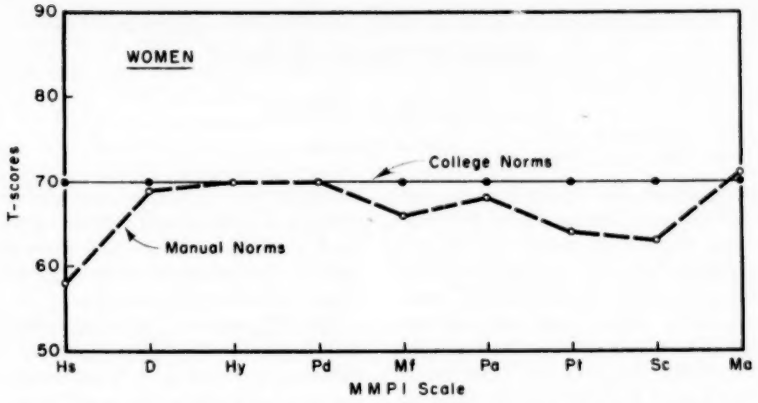


FIG. 2. T-scores on the Manual norms of raw scores corresponding to T-scores of 70 on the College norms (Women).

has T-values of 69 and 55 on the manual and college norms respectively.

3. Raw scores on the remaining scales had very similar T-values on both sets of norms, especially at that part of the scale, the upper end, which is claimed to have clinical significance.

Another method of comparing the two sets of norms was suggested. Raw scores corresponding to T-scores of 70 on the college norms were selected as a basis of comparison. The T-score equivalents on the manual norms were obtained for these same raw scores and are graphed in Figures 1 and 2. A raw score of 70 was chosen because Hathaway and McKinley consider it to be "a borderline score, although useful interpretation will always depend upon the clinician's experience with a given group" (**1**, p. 8).

At the upper ends of the scales, it is apparent that discrepancies between college and manual norms are found on the Hs, D and Mf scales for men, and on Hs, Pt and Sc for women.

In general, the differences between the two sets of norms appear to be relatively minor. However, multivariate analysis by means of Hotelling's T (**5**) might reveal significant differences between the standardization and college groups.

Received October 29, 1952.

References

1. Hathaway, S. R. and McKinley, J. C. *The Minnesota Multiphasic Personality Inventory.* New York: The Psychological Corporation, 1943.
2. McKinley, J. C., Hathaway, S. R., and Meehl, P. E. The Minnesota Multiphasic Personality Inventory: VI. The K Scale. *J. cons. Psychol.*, 1948, **12**, 20–31.
3. Nance, R. D. Masculinity-femininity in prospective teachers. *J. educ. Res.*, 1949, **42**, 658–666.
4. Tyler, F. T. A factorial analysis of fifteen MMPI scales. *J. cons. Psychol.*, 1951, **15**, 451–456.
5. Tyler, F. T. Some examples of multivariate analysis in educational and psychological research. *Psychometrika*, 1952, **17**, 289–296.
6. Tyler, F. T. and Michaelis, J. U. *University-student norms for the Minnesota Multiphasic Personality Inventory.* Mimeographed. Available on request to the writers.
7. Tyler, F. T. and Michaelis, J. U. K-scores applied to MMPI scales for college women. (To appear in *Educ. psychol. Measmt.*)

THE JOURNAL OF APPLIED PSYCHOLOGY
Vol. 37, No. 4, 1953

Socio-economic Status and Culturally-weighted Test Scores of Negro Subjects

Frank C. J. McGurk

Lehigh University

A previous article (6) reported on the non-cultural and cultural test scores of 213 pairs of white and Negro high school seniors who had been matched for age, school attendance, school curriculum, and eleven selected socio-economic factors. All subjects were between the ages of sixteen and twenty; the mean age for the whites was 18.1, and the mean age for the Negroes was 18.2. The subjects were obtained from schools in southeastern Pennsylvania and northern New Jersey.

Socio-economic status was defined in terms of the score obtained on a revision of the Sims Record Card. The high Negro socio-economic group is composed of those Negro subjects whose revised Sims scores were in the highest 25% of the range of Negro Sims scores. The low Negro socio-economic group is composed of those Negro subjects whose revised Sims scores were in the lowest 25% of the range of Negro Sims scores.

Test questions were defined as non-cultural and cultural according to the pooled judgments of 78 school teachers, psychologists, and sociologists. Non-cultural questions are those which the judges considered least culturally-weighted; cultural questions are those considered heavily weighted with cultural experiences.

Complete details on the selection and matching of subjects, the revision of the Sims Record Card and the determination of socio-economic status, and the dichotomizing of the questions can be found in the previous article (6).

This paper is concerned with the problem of how the socio-economic status of Negro subjects affects their differential test performance on non-culturally-weighted and culturally-weighted test questions.

Results

Table 1 shows the mean non-cultural and cultural scores of the high and low socio-economic groups. Since the standard deviations of the non-cultural and cultural scores are almost identical (4.62 for the non-cultural and 4.64 for the cultural) the findings will be presented in terms of raw scores.

The difference in mean non-cultural score that is associated with a difference in socio-economic level (the H − L difference) is 2.22; the H − L difference for the cultural questions is 1.21. Thus, in the comparison between the highest and lowest socio-economic group of Negroes, a greater difference is obtained on the non-cultural, ***not on the cultural,*** questions. The difference between the two H − L differences is significant only at the 20% level.

Table 1

Negro Non-cultural and Cultural Mean Raw Scores by Socio-economic Groups

Type of Question	Socio-economic Group*		Difference (H−L)
	High	Low	
Non-cultural	13.62	11.40	2.22
Cultural	9.81	8.60	1.21
Net change			1.01
SE of net change			0.89
t			1.24
P			20% approx.

* N for each group = 53.

Discussion

In his study of the effects of length of residence in New York City on Negro test scores, Klineberg found that, as the length of residence increased, test scores increased. He also found that his results ". . . are much clearer for the linguistic tests than for the performance tests" (5, p. 44).

Various writers have described linguistic or verbal tests as sensitive to differences in socio-

economic status because of the culturally-weighted content of such tests (2, 3, 4). Others have described performance or non-verbal tests as not being as sensitive to socio-economic level because the content of these tests is not so culturally-weighted (1, 3, 5). Hence, Klineberg's findings have been interpreted to mean that increasing the socio-economic status of the Negro should be accompanied by a greater improvement on culturally-weighted material than on the less culturally-weighted material.

The present findings do not support such an interpretation. The present data show that the test superiority of the Negro of high socio-economic status over the Negro of low socio-economic status is associated more with a superior performance on the non-cultural questions than on the cultural questions.

Received October 6, 1952.

References

1. Alpers, T. G. and Boring, E. G. Intelligence test scores of northern and southern white and Negro recruits in 1918. *J. abnorm. soc. Psychol.*, 1944, **39**, 471–474.
2. Bean, K. L. Negro responses to certain intelligence test items. *J. Psychol.*, 1941, **12**, 191–198.
3. Bean, K. L. Negro responses to verbal and non-verbal test material. *J. Psychol.*, 1942, **13**, 343–353.
4. Brown, F. An experimental and critical study of the intelligence of Negro and white kindergarten children. *J. genet. Psychol.*, 1944, **65**, 161–175.
5. Klineberg, O. Tests of Negro intelligence. In Otto Klineberg (Ed.), *Characteristics of the American Negro*. New York: Harper & Bros., 1944.
6. McGurk, F. C. J. *Comparison of the performance of Negro and white high school seniors on cultural and non-cultural psychological test questions*. Washington: The Catholic University of America Press, 1951 (microcard).

THE JOURNAL OF APPLIED PSYCHOLOGY
Vol. 37, No. 4, 1953

The Relationship Between Rater Characteristics and Validity of Ratings

Dorothy E. Schneider and A. G. Bayroff [1]

Personnel Research Section, Personnel Research and Procedures Branch, TAGO, Department of the Army, Washington, D. C.

The Personnel Research Section, Personnel Research and Procedures Branch, TAGO has been conducting a series of studies aimed at identifying personal characteristics of raters which are associated with more valid ratings. The problem is of particular significance to this office which is responsible for the development of efficiency reports for Army-wide use. It is also important because the absence of more objective criteria makes it necessary to use multiple ratings in the validation of other types of instruments. If personal characteristics of the rater which are related to the validity of ratings he gives could be found, it might be possible to control them, perhaps in the selection of raters, or in the form of a "correction score." While this may not be feasible on official efficiency ratings, it could profitably be done when obtaining ratings for use as criteria. Further, it may be possible to develop new techniques which would be more independent of the personal characteristics of the rater.

Problem

The study reported here was concerned with the validity of ratings by raters differing in three characteristics: aptitude test score, academic achievement, and rated over-all value to the Army.

Method

Subjects. The population consisted of 400 officers (primarily majors and lieutenant colonels) enrolled as students at the Army Command and General Staff College. The objective of this college is to train potential division commanders and general staff officers, and its students represent a highly selected group. The course was 42 weeks long and the students were in close contact with each other during the entire period.

The officers were assembled in groups of 33–40, the size of the classes, and asked to rate their class associates. In this manner each officer served as both rater and ratee.

Design of the Study. This report covers one aspect of a larger research program on rating methodology. Only those aspects of the design pertinent to the present study will be presented here.

The Criterion. The criterion measure was an appraisal of over-all value to the Army. The total population was randomly split into two groups, A and B, each consisting of 200 officers. Officers in Group A provided the criterion rankings for all officers in the population. Each member of this group was required to evaluate his class associates by placing them in the highest, middle, or lowest third of the class. Each rater evaluated 20 ratees from Group A and 20 ratees from Group B; thus each member of the total population was evaluated by 20 raters. The criterion score was the mean of the 20 rankings.

Anonymous Rating Scale. This was an 8-point rating scale on which the officers were rated on over-all value to the Army. Each of the scale units was described in some detail. Group A completed 20 such ratings, 10 on officers in Group A, and 10 on officers in Group B. Each of the 400 officers thus received 10 ratings with this instrument, and the mean of these ratings was used as the score. The raters were informed that these were to be held in confidence, that the ratings would not be made available to the College or the Army, and that the raters need not sign their names. These ratings will be referred to hereafter as the unsigned ratings.

Identified Official Rating Scale. This was the same 8-point rating scale just described. It was used by Group B, each officer rating

[1] The authors wish to acknowledge the assistance of various members of the staff, particularly Dr. E. A. Rundquist, Mr. A. H. Birnbaum, and Mr. Joel T. Campbell. The opinions expressed in this paper are those of the authors and do not necessarily reflect those of the Department of the Army.

10 ratees from Group A and 10 ratees from Group B. Each of the 400 officers thus received 10 such ratings, and the mean rating served as his score. The raters were informed that these ratings would be available for official use for one year, and they were required to sign the ratings. These ratings will be referred to hereafter as the signed ratings.

Forced Choice Pairs. The forced choice technique has been used in Army officer efficiency reports from July 1947 to September 1950. Its use in efficiency reports was discontinued largely because the form of forced choice used was not acceptable. For an extended discussion, see Baier.[2] Research has continued, nevertheless, with the objective of developing more acceptable forms. One of these forms contained the grouping of phrases in pairs instead of tetrads as formerly used. The phrases used were taken directly from the former Efficiency Report, Form 67–1. Twenty-four pairs were employed in this study, members of each apparently equally favorable, but differing in their ability to discriminate better from poor officers as determined by their criterion scores. The rater checked the phrase of each pair which he felt was **More Descriptive** of the ratee. Scoring was based on the discrimination values of the phrases checked. One such rating form was completed for each officer in the population by the raters of Group B.

Controlled Check List. The controlled check list was the second form of the forced choice technique included in this study. The phrases used in the 24 forced choice pairs were grouped into two sets of 24 phrases each. The rater selected the 12 phrases of each set which he felt were most descriptive of the ratee and scoring was based on the discrimination values of the phrases checked. One controlled check list was rendered on each officer in the population by the raters of Group A.

Rater Variables. Raters were divided into equal thirds, highest, middle, and lowest, on the basis of scores on each of the variables described below.

[2] Baier, D. E. Reply to Travers' "A critical review of the validity and rationale of the forced-choice technique." *Psychol. Bull.*, 1951, **48**, 421–434.

Aptitude. Scores on the Officer Classification Test (OCT), a high level aptitude test, were available for all officers.

Final Class Standing. Officer students are ranked on academic work periodically during their attendance. The last ranking is the Final Class Standing (FCS). Prior to this study, class standing and the other variables were transferred to IBM cards in such manner that the identification of individual officers was no longer possible.

Derived Final Class Standing. To remove the contribution of aptitude from Final Class Standing, predicted FCS was obtained from the regression of FCS on OCT. The difference between actual and predicted FCS of each officer was used as another measure of achievement referred to here as the Derived Final Class Standing (Derived FCS).

Over-all Value. This score was the mean of the 20 criterion rankings received by each rater.

After the raters were divided into thirds, the validity of ratings by each of these thirds on each of the rating variables was obtained. For the 8-point scales, multiple ratings per ratee were available and validities were obtained for both single ratings and mean of ratings given by each rater. For the two forced choice forms, validities were obtained for single ratings only.

Results

The validities of ratings given by raters of each of the levels of rater characteristics studied are presented in Table 1.

In general, a direct relationship existed between the validity of the ratings given and measures of the raters' over-all value, aptitude, and achievement. Except in a few instances, validity coefficients for the highest thirds were higher than those for the middle thirds and those for the middle thirds were higher than the validity coefficients for the lowest thirds. In no case did ratings by raters in the lowest thirds have higher validities than ratings by raters in the highest thirds.

It will be noted that although the validities of ratings with the 8-point scale were slightly higher than those with the other types of ratings, the lower validity by poorer officers was

Table 1

Validity of Ratings by Raters of Different Levels of Ability and Achievement

Raters Divided into Highest, Middle, and Lowest Thirds on		Validity of Ratings on					
		Unsigned 8-Point Rating		Signed 8-Point Rating		Forced Choice Pairs	Controlled Check List
		Single Ratings	Mean Ratings	Single Ratings	Mean Ratings	Single Ratings	Single Ratings
Aptitude (OCT)	H	.57	.74	.50	.72	.46	.45
	M	.50	.71	.51	.66	.48	.43
	L	.51	.69	.48	.60	.30	.44
Final Class Standing (FCS)	H	.58	.74	.54	.73	.50	.50
	M	.54	.71	.49	.65	.31	.46
	L	.47	.60	.44	.57	.40	.37
Derived Final Class Standing (Derived FCS)	H	.58	.74	.54	.72	.49	.55
	M	.55	.75	.50	.61	.36	.36
	L	.46	.64	.42	.62	.35	.42
Criterion Ranking	H	.61	.79	.55	.69	.49	.48
	M	.48	.66	.51	.66	.30	.48
	L	.49	.66	.45	.65	.43	.31

present in all types of scales. Of the two 8-point rating scales, the validities were somewhat higher for the unsigned than for the signed. However, Group A, which rendered the unsigned ratings, also performed the criterion ranking, so some degree of rater contamination may be present.

Both forced choice forms had greater decreases in validity from the highest third to the middle third of the rater groups than did the 8-point scales. There were three decreases as large as .19 for the forced choice forms, and the only comparably large decrease for the 8-point scales was .13. It thus appeared that raters of lesser ability, as defined in this study, produced more accurate ratings with the conventional technique than with forced choice or controlled check lists.

Summary

Officers at the Army Command and General Staff College rated each other using four techniques: two 8-point scales of over-all value (one signed by the rater and the other unsigned), and two forms of the forced choice technique (forced choice pairs and a controlled check list). Rater groups were divided into highest, middle, and lowest thirds on the basis of aptitude test score, final class standing, final class standing predicted from aptitude test score, and on the criterion rank of over-all value achieved. For each third of the groups, separate validity estimates of ratings made were computed for both individual ratings and mean ratings where available.

It was found that raters who scored high on aptitude, achievement at the College, and over-all value to the Army produced more valid ratings than did raters who scored lower on these variables. This trend was highly consistent for the 8-point rating scales, and clear, though not as direct, for forced choice pairs and the controlled check list.

Received October 15, 1952.

THE JOURNAL OF APPLIED PSYCHOLOGY
Vol. 37, No. 4, 1953

The Actuality Measure in the Study of Public Opinion

Peter R. Hofstaetter
The Catholic University of America

When he first suggested an "actuality-measure" (A) this author (7, 8, 9) aimed at combining two aspects of rather strongly ego-involved states of public opinion. He hypothesized that a population of respondents shows relatively great ego-involvement in a question if two conditions are fulfilled: (a) Few "don't know" responses; and (b) An even split between the Yes- and No-responses.[1]

The formula which purports to measure the actuality of a given question for a given population of respondents was based on these two notions:

$$A = \sqrt{\frac{p_+ \cdot p_-}{p_0^2}}.$$

The evidence this writer has meanwhile collected seems to confirm the assumptions underlying this formula (10).

In this paper we shall first try to summarize the available evidence for the designation of A as a measure of "actuality" or of the involvement of a group in a topic. This is supposed to furnish an empirical check of the validity of our measure.[2] Our second attempt will then be to present a statistical model from which the A-measure can be derived and which allows us to account for the empirical properties of A.

After having rather regularly analyzed poll-reports during the last few years at least five conditions became clear which tend to result into relatively high actuality of questions. Each statement will be illustrated by one example; additional examples could easily be obtained.

[1] We shall concern ourselves only with questions that are answered in terms of a trichotomy, i.e., in the affirmative ("Yes"), in the negative ("No") or undecided ("Don't know"). The respective percentages of the responses falling in each of these categories are referred to as p_+, p_- and p_0.

[2] A better check would obviously consist of an attempt to predict the relative values of A for questions which had been answered by different populations. Work in this direction is underway but to report on it would go beyond the scope of the present paper.

1. The more pertinent an issue is felt to be the higher is its actuality for a particular group of respondents. See Table 1.

Table 1

Distribution of Answers to the Question: "Do you think cigarette smoking is harmful or not?"†

Respondents	Harmful %	Don't know %	Not Harmful %	A
Cigarette smokers	52	3	45	16.12
Non-cigarette smokers	66	10	24	3.98

† AIPO., Dec., 1949.

2. The more imminent the event is to which the question refers the higher is the actuality of the question. See Table 2.

3. The greater the change involved in the question the higher is the actuality. See Table 3.

4. The richer the available background of experience the higher is the actuality of the question (Saenger and Gordon, 19). See Table 4.

5. The higher the educational and/or the socio-economic level of the respondents the higher actualities questions tend to attain. The data in Table 5 are based on 14 questions reported by the Psychological Corporation's Barometer (Link, 13, 14, 15, 16).

One may infer from these data that quite a few of the usual poll questions are tuned to the mentality of the above-average strata of our population. Thus they often reflect the problems of the "intelligentsia" rather than those faced by the "man on the street." There are, however, interesting exceptions to the general trend (Link, 13). See Table 6.

Cases like the one just mentioned can probably be understood in terms of point one of the present list, i.e., with reference to the greater pertinence of the question to the lower socio-economic levels.

Table 2

Distribution of Answers to the Question: "Do you think the U. S. will find itself in another world war within, say, the next year? (five years?)"†

Date of Interview	Within One Year				Within Five Years			
	Yes %	No Op. %	No %	A	Yes %	No Op. %	No %	A
May 1950	22	8	70	4.91	57	19	24	2.45
Sept. 1950	29	16	55	2.50	58	22	20	1.55
July 1951	26	10	64	4.08	56	26	18	1.22
Dec. 1952	21	12	67	3.13	48	27	25	1.28

† AIPO.

Table 3

Distribution of Answers to the Question: "Would you approve or disapprove of the following changes of postage which have been suggested?"†

Change of Postage	Per Cent Increase	Approve %	No Op. %	Disapprove %	A
Postcards from 1 ct. to 2 cts.	100	52	7	41	6.59
Reg. mail from 3 cts. to 4 cts.	33	33	8	59	5.51
Air mail from 6 cts. to 7 cts.	17	51	9	40	5.02

† AIPO., Aug., 1949.

Taken together, these examples seem to suggest the empirical validity of the actuality-measure. They represent, however, only a small sample from the total available evidence.

While the empirical validity of the actuality-measure seems to stand on relatively firm ground the arbitrariness of its derivation caused the present writer considerable headache. It will be the task of the present discussion to remedy this shortcoming.

Let us assume, for the sake of a model, that our respondents were to give their answers on the basis of two balls drawn from two urns. Each of these urns contains "a" per cent white and "b" per cent black balls. The following combinations can thus occur: two white balls, one white and one black ball, two black balls. According to the rules of our game the respondents will have to render a "Yes-statement" whenever they draw two white balls, and a "No-statement" if both balls are black. In case they draw one white and one black ball the answer is going to be "Undecided," or "Don't know," "Uncertain," "No opinion," etc.

The frequencies with which we expect these

Table 4

Distribution of Answers to the Question: "Do you believe in the efficiency of the New York State law against discrimination?"

Respondents	Yes %	Don't Know %	No %	A
Persons who had experienced job discrimination	15	14	71	2.33
Others	27	21	52	1.78

Table 5

Actuality as a Function of the Socio-economic Level of the Respondents

Socio-economic Level	Average A	Average A (%) (Total = 100)
A (highest 10 per cent)	4.91	151
B (next 30 per cent)	4.04	131
C (next 40 per cent)	3.28	100
D (lowest 20 per cent)	2.03	61
Total	3.32	100

Table 6

Distribution of Answers to the Question: "Have businessmen a right to shut down?"

Socio-economic Level	Yes %	Uncertain %	No %	A	A %
A	60	8	32	5.48	83
B	57	7	36	6.48	98
C	46	7	47	6.64	100
D	41	6	53	7.78	117
Total	49.5	8	43.5	6.63	100

Table 7

Distribution of Answers to the Question: "Do you think that Socialism in England will succeed or fail?"

Socio-economic Level	Succeed %	Don't Know %	Fail %	A	A %
A	19	24	57	1.37	236
B	13	31	56	0.87	150
C	10	42	48	0.52	89
D	11	60	29	0.30	52
Total	12	41	47	0.58	100

responses to occur will follow the binomial expansion:

$$(a + b)^2 = a^2 \text{ (Yes)} + 2ab \text{ (Don't know)} + b^2 \text{ (No)}.$$

The middle term in this expansion becomes thus:

$$2ab = 2\sqrt{a^2b^2} = p_0.$$

We replace now a^2 and b^2 by the observed percentages of the Yes- and No-responses (p_+ and p_- respectively) and thus obtain for the middle term:

$$p_0 = 2\sqrt{p_+p_-}.$$

For all binomial distributions the following relationship holds:

$$P = \frac{2\sqrt{p_+p_-}}{p_0} = 1.00.$$

This provides us with an easy way to ascertain whether or not an empirical distribution follows the binomial pattern. We have to reject the simple binomial (or "Bernoullian") model whenever the ratio P differs significantly from 1.00.

The magnitude P becomes thus a measure for the applicability of the Bernoullian model to an observed distribution of responses. It can, however, be shown that our actuality-measure (A) serves this very same purpose equally well:

$$P^2 = \frac{4p_+p_-}{p_0^2} = 4A^2; \text{ consequently: } A = \tfrac{1}{2}P.$$

Since P equals 1.00 for all binomial distributions A necessarily equals 0.50.

It thus becomes clear that our actuality-measure is nothing but a quantitative expression for the appropriateness of the Bernoullian model for the representation of an empirical distribution of responses. It should be noted that this model does not presuppose any specific proportionality between the two initial probabilities "a" and "b." The appearance of a rather substantial majority for either "Yes" or "No" is therefore compatible with the Bernoullian model.

The examples given in the first part of this paper have shown rather consistently actualities higher than 0.50. Indeed, it is not easy to locate questions which yield actualities that low or lower. An example may be drawn, however, from one of Link's polls (**14**). See Table 7.

The response distribution of Group C ($A = 0.52$) comes indeed very close to the expansion of $(0.30 + 0.70)^2 = 9\%$ (Succeed) + 42% (Don't know) + 49% (Fail).

Since poll data conform only rarely to this pattern we have to recognize that the Bernoullian model is too simple to account for them. But we know already that its failure in that respect can be gauged from either P or A.

The next step we have to undertake is to replace the model of two urns with the same probabilities ("a" and "b") in each by the model of two urns with different probabilities. Obviously, our first model can be considered as a special case of this second, more general model.

Let us assume that the respondents base their decisions on the drawing of one ball from each of two urns where the following

Table 8

The Markov Model of a Response Distribution

Urn	Probabilities			A
I	$a_1=0.50$		$b_1=0.50$	
II	$a_2=0.70$	$b_2=0.30$		
III		$a_3=0.10$	$b_3=0.90$	
Responses				
%	$p_+=35$	$p_0=20$	$p_-=45$	1.98

probabilities prevail:

$$a_1 \neq a_2 \text{ and } b_1 \neq b_2.$$

The corresponding percentages of the three categories of answers are:

$$p_+ = a_1 \cdot a_2$$
$$p_0 = (a_1 \cdot b_2) + (a_2 \cdot b_1)$$
$$p_- = b_1 \cdot b_2$$

This model always results in distributions which have actualities of less than 0.50. Its usefulness is therefore drastically limited.

This will no longer be the case with the third model which employs the statistical theory of the "Markov Chains" (5). Let us assume that there are three urns with different probabilities for white and black balls:

$$a_2 > a_1 > a_3 \text{ and } b_2 < b_1 < b_3.$$

Whenever a subject has drawn a white ball from Urn I he is bound to draw his second ball from Urn II where the probability of a white ball is even greater than in the first urn. In the event of a black ball being drawn from the first urn the subject is required to draw the second ball from Urn III which gives him again an even higher probability to draw a black ball than the first urn did. This model abandons the independence condition of the Bernoullian model. A numerical example may help to clarify this notion. See Table 8.

The essential feature of this model is that the probabilities which prevail for the drawing of the second ball depend on the result of the first drawing. Such probabilities have been called "dependent" or "conditional" probabilities. It can be seen easily that the Markov model includes the Bernoullian model as a special case of $a_1 = a_2 = a_3$. The outcome of a so-called Markov process is defined by the following equations:

$$p_+ = a_1 \cdot a_2$$
$$p_0 = (a_1 \cdot b_2) + (b_1 \cdot a_3)$$
$$p_- = b_1 \cdot b_3$$

Though it may seem as if we had three equations for the determination of the three unknown variables (a_1, a_2 and a_3) we will not be able to solve these equations since only two of the p-values are independent. The problem is therefore bound to remain underdetermined unless we introduce an arbitrary assumption with respect to the initial probability a_1. Our assumption may, for instance, be: $a_1 = b_1 = 0.50$.[3] Consequently, the following equations will define the conditional probabilities:

$$a_2 = \frac{1}{a_1} p_+ = 2p_+;$$

$$a_3 = \frac{p_0 - a_1 b_2}{b_1} = 2(p_+ + p_0) - 1.00.$$

Since the observational data we collect by means of poll questions do not allow us to determine all three probabilities needed for the Markov model its usefulness may seem to be sort of academic. Yet, it should be pointed out that this model shows one property which furthers our understanding of the processes which underlie the formation of public opinion. In order to become aware of this aspect we have only to replace the term "urn" by the term "source of information." The Markov model tells us that the conclusions our respondents draw from one source of information (these conclusions may be either favorable or unfavorable with respect to the question at issue) determine their exposure to other sources of information (some of them being more and others less likely to confirm a favorable attitude). What accounts ultimately for an observed response distribution is thus a certain amount of selectivity with respect to the sources of information to which the respondents expose themselves. What this boils down to is that the drawing from different sources of information does not occur in a statistically independent manner. On

[3] This assumption becomes untenable in the case of: $p_+ + p_0 < 50$.

the contrary, the sources of information which our respondents actually recognize are in most cases intercorrelated.

Once the first choice has occurred the subsequent choices tend to be made on the basis of sources of information which are more likely to support the initial choice than to contradict it. Thus a series of choices gets underway which assures the individual subject of a relatively stable and un-ambivalent attitude. By the same token we may infer that neutral responses ("Don't know," etc.) tend to vanish in the group of respondents as a whole. What we have called "actuality" and have operationally defined by our formula thus serves as an indicator for the amount of "bundling" of sources of information which prevails in a given group of respondents with respect to a given issue.

Table 9

The Markov Model of Two Correlated Sources

Before Debate					After Debate			
p_+	p_0	p_-	A		p_+	p_0	p_-	A
36	46	18	0.55		36	18	46	2.26
		Urn II				Urn II		
Urn I		a_2	b_2	Sum		a_2	b_2	Sum
a_1		36	23	59		36	9	45
b_1		23	18	41		9	46	55
Sum		59	41	100		45	55	100
		$r_t = 0.08$				$r_t = 0.83$		

In order to explore this last point further we can restate the Markov model in the following way: We assume two sources of information ("urns") with the same probabilities for either positive or negative statements in each of them: $a_1 = a_2$; $b_1 = b_2$. We can thus set up a fourfold table in which all sorts of trichotomic response distributions can be arranged as is shown in Table 9.[4]

The data of Table 9 are taken from an experiment by Millson (17) in which the author arranged for group discussions on the topic of unemployment insurance. Before and again after debate the participants were asked to state their opinions. As can be seen from Table 9 the debate resulted in a heightened actuality of the question. This is a fairly common event (10). In the lower half of Table 9 the same data have been brought in the form of fourfold tables in order to allow for the computation of tetrachoric correlations. The value of p_0 had thus to be divided into two cells, i.e., the combination of a_1 with b_2 and of a_2 with b_1. One half of p_0 was allotted to either cell.

The interpretation that can be given to these data reads: Whereas the respondents used the sources of information available to them before the debate in an (almost) independent manner a considerable degree of bundling or patterning occurred after the debate. The rise in actuality (from 0.55 to 2.26) corresponds with an increase in the degree of bundling (from $r_t = 0.08$ to $r_t = 0.83$). In fact, there exists, in general, a definite relationship between A and the corresponding coefficient of correlation r_t. See Figure 1.[5]

Actualities below 0.50 correspond to negative correlations. This means that the respondents who have drawn their conclusions from the first source of information in a positive manner are more likely to draw a negative conclusion from the second source they consult, and vice versa. This is seldom the case. An example can be seen in the behavior of socio-economic group D in Table 7. In introspective terms this may be either considered as "disappointment" or as a striving for objectivity.

As we reach the end of our present investi-

[4] The tetrachoric correlation coefficients (r_t) are read from the well-known Thurstone Tables (3).

[5] An empirical equation has been fitted to this curve:

$$r_t = 0.96 - 1.78\ e^{-1.338\ A}$$

For A-values over 0.50 the agreement between expected and observed r_t-values is excellent, below 0.50 it is tolerable. It had been hoped that this relationship would provide us with a test of significance for A. This hope may, however, not materialize because of the great difficulties that enter into the determination of the significance of differences between tetrachoric coefficients. In the future a significance test for A in terms of the corresponding Chi-squares seems to be more likely. It can, however, be reasonably inferred from Figure 1 that relatively small numerical differences between A-values in the range from 0.50 through 3.50 are much more likely to be significant than numerically larger differences between higher A-values.

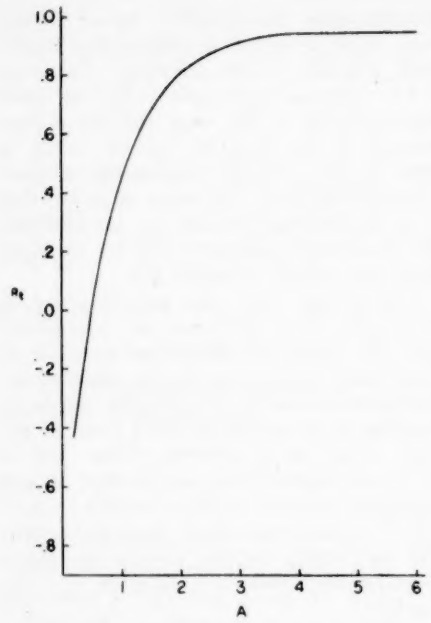


FIG. 1. The relationship between actuality (A) and the bundling of sources of information (r_t).

gation it may be said that the actuality-measure indicates the amount of discrepancy that exists between an observed response distribution and the kind of distribution that one would expect on the basis of the Bernoullian model. This discrepancy can be expressed in terms of the correlated functioning of two sources of information (Markov model).

Needless to say, no model can ever be taken literally. It is quite artificial to argue that always two and just two sources enter into the determination of a person's attitude towards a given social issue. Indeed, there is little doubt that we are exposed to many more sources of information with respect to most issues.[6] It seems, however, that the more sources of information the ordinary person, in accordance with his socio-economic and/or educational level, consults, the stronger becomes his chance to hit upon correlated sources. This is in line with point five in the first part of this paper. The other four points previously made can now be summarized by saying that the more important a decision we have to make, the more prone we are to expose ourselves to correlated sources of information. Yet, this statement requires at least one qualification. As has been shown in an earlier paper (9) there exists a negative correlation between the actuality and the experienced difficulty of questions. Our subjective experience that a question is difficult seems to correspond with an un-bundling of sources of information. Ultimately this causes us to shift in the direction of the Bernoullian model.

To recognize fully the eventually large number of sources of information would, of course, complicate our derivations. The basic postulate, however, will probably remain unaltered, that is, that one can either use the multitude of sources of information in a complete and unselected manner or use them selectively. From a theoretical point of view the first procedure is undoubtedly preferable; in practice the latter seems to be the rule. To follow this "rule" minimizes the ambiguity of perceived situations and thus eases the burden of uncertainty.[7] *Ceteris paribus* this tendency may be expected to show up more strongly the more deeply ego-involved we are in certain issues. Actuality measures the strength of this tendency.

Summary

1. Poll questions which can be answered by either "Yes," "No" or "Don't know" yield response distributions which can be compared

[6] The "interference-hypothesis" which asserts "that any communication which successfully modifies a person's beliefs will reduce the opinion-impact of any subsequent event or communication that tends to produce antithetical beliefs" has been recently confirmed (11).

[7] This very same mechanism has been described in several situations by different terms. For instance as the "selectivity" of learning and forgetting (5, 12), or as "intolerance of ambiguity" (6) with regard to perception. The relationship between "intolerance of ambiguity" and ethnocentrism has been demonstrated (1). A comparable tendency to escape from ambiguity seems to underlie the "Halo-Effect" (Thorndike) in personality judgment. The correlated use of indicators refers also to this situation (2). H. Werner's (20) concept of "rigidity" is also based upon the notion of either too much isolation or too much overlapping of subareas, i.e., upon their positive or negative correlation. The "hypothesis-confirmation theory" of Postman (18) provides a suitable frame of reference for these phenomena.

with binomial expansions.

2. The actuality-measure indicates the discrepancy between an observed response distribution and that expected from the binomial model.

3. All kinds of trichotomic response distributions can be fitted to the Markov model of dependent probabilities.

4. Under simplifying conditions, the actuality-measure becomes an indicator of the amount of bundling (or correlation) among the sources of information to which the respondents have exposed themselves.

5. The bundling of sources of information increases with the subjectively felt importance of the issue to which the question relates and also—in general—with the socio-economic and/or the educational level of the respondents. It decreases with the subjectively felt difficulty of the issue.

Received August 20, 1952.

References

1. Block, J. and Block, Jeanne. An investigation of the relationship between intolerance of ambiguity and ethnocentrism. *J. Pers.*, 1951, **19**, 303–311.
2. Brunswik, E. *Systematic and representative design of experiments.* Berkeley: Univ. of Calif. Press, 1947.
3. Chesire, L., Saffir, M., and Thurstone, L. L. *Computing diagrams for the tetrachoric correlation coefficient.* Chicago: Univ. of Chicago Bookstore, 1933.
4. Edwards, A. L. Political frames of reference as a factor influencing recognition. *J. abnorm. soc. Psychol.*, 1941, **36**, 34–61.
5. Feller, W. *An introduction to probability theory and its applications.* New York: Wiley, 1950.
6. Frenkel-Brunswik, E. Intolerance of ambiguity as an emotional and perceptual personality variable. *J. Pers.*, 1949, **18**, 108–143.
7. Hofstaetter, P. R. *Die Psychologie der oeffentlichen Meinung.* Wien: Braumueller, 1949.
8. Hofstaetter, P. R. The actuality of questions. *Int. J. Opin. Attitude Res.*, 1950, **4**, 16–26.
9. Hofstaetter, P. R. Importance and actuality. *Int. J. Opin. Attitude Res.*, 1951, **5**, 31–52.
10. Hofstaetter, P. R. *Einfuehrung in die Sozialpsychologie.* Wien: Humboldt, 1953.
11. Janis, I. L., Lumsdaine, A. A., and Gladstone, A. I. Effects of preparatory communications on reactions to a subsequent news event. *Publ. Opin. Quart.*, 1951, **15**, 487–518.
12. Levine, J. M., and Murphy, G. The learning and forgetting of controversial material. *J. abnorm. soc. Psychol.*, 1943, **38**, 507–517.
13. Link, H. C. The Psychological Corporation's index of public opinion. *J. appl. Psychol.*, 1946, **30**, 297–309.
14. Link, H. C. The ninety-fourth issue of the Psychological Barometer and a note on its fifteenth anniversary. *J. appl. Psychol.*, 1948, **32**, 105–117.
15. Link, H. C., and Freiberg, A. D. The ninety-seventh Psychological Barometer. *J. appl. Psychol.*, 1948, **32**, 443–451.
16. Link, H. C., and Freiberg, A. D. The Psychological Barometer on Communism, Americanism and Socialism. *J. appl. Psychol.*, 1949, **33**, 6–14.
17. Millson, W. A. D. Problems in measuring audience reaction. *Quart. J. Speech*, 1932, **18**, 621–637.
18. Postman, L. Toward a general theory of cognition. In J. H. Rohrer and M. Sherif (Eds.), *Social psychology at the crossroad.* New York: Harper, 1951.
19. Saenger, G., and Gordon, N. S. The influence of discrimination on minority group members in its relation to attempts to combat discrimination. *J. soc. Psychol.*, 1950, **31**, 95–120.
20. Werner, H. The concept of rigidity; a critical evaluation. *Psychol. Rev.*, 1946, **43**, 43–52.

THE JOURNAL OF APPLIED PSYCHOLOGY
Vol. 37, No. 4, 1953

A Biasing Factor in Essay Response Frequency

Erwin K. Taylor and Dorothy E. Schneider

Personnel Research Institute, Western Reserve University

In order to get descriptive phrases for use in developing an evaluation form, questionnaires were mailed to 955 randomly selected members of a professional organization, the American Dietetic Association. The questionnaire explained the purpose of the study and asked each respondent to describe an associate at a level of competence specified on the form. Each questionnaire specified a single level out of ten possible levels. Ninety-five or ninety-six questionnaires were sent out for each level. Of the ten levels, "1" was considered best, and "10" poorest. Although there was some feeling on the part of the present writers that the returns, being on an anonymous and voluntary basis, would yield either a J- or U-shaped distribution, there was no published evidence to justify unequal distribution in mailing of the forms. Hence, each level was requested from an equal number of recipients of the questionnaire.

In addition to the descriptive essays, respondents were asked a number of questions regarding the type of work engaged in when they were acquainted with the dietitian described, length of acquaintance, and working relationship between the two.

The questionnaires were mailed from Cleveland on June 19, 1952; the first reply was received on June 24, and by July 10, the "deadline" requested on the questionnaire form, 130 forms had been returned. Within the two weeks following, 14 more were received, bringing the total to 144, or about a 15% return.

The forms which were returned reveal some interesting points. The distribution at each level of competence is shown in Table 1, according to whether the respondent selected a subordinate, co-worker, superior, or student as subject for the essay.

It will first be noted that the distribution, while not horizontal, does not reveal sufficient U or J shape to be so named. The mean of 5.4 is very close to the 5.5 which would be expected in a horizontal distribution. At only two levels, 4 and 6, are the frequencies more than 5 below the maximum frequency.

A Chi-square test was applied to the total frequencies and revealed no statistically significant deviation from the theoretical recti-

Table 1

Distribution of Returns on Questionnaire Form Requesting Descriptive Essays

Level of Competence of Person Described	Relation of Subject to Writer of Essay				
	Subordinate	Co-Worker	Superior	Student	Total
1 (Best)	6	0	10	2	18
2	5	3	10	0	18
3	5	4	4	1	14
4	1	2	5	2	10
5	4	5	7	0	16
6	4	3	2	0	9
7	3	6	3	1	13
8	9	2	2	1	14
9	4	4	4	2	14
10 (Poorest)	9	5	4	0	18
Total	50	34	51	9	144
Mean	5.96	6.26	4.35	5.11	5.40

linear distribution. This indicates the absence of bias in the returns in terms of the level of competence of the individual to be described.

Thus, while it has been conjectured that it is easier to evaluate personnel at the extremes of the effectiveness distribution, it does not appear to be necessary to take this factor into account in planning the collection of data of this nature. Responses were received with almost equal frequency for ratees at each of the ten effectiveness levels used in this study.

When no restriction was placed upon the selection of subjects for the essays, other than the level of competence, about the same number of supervisors and subordinates were selected. However, some rather striking differences will be observed in the mean levels of competence of the two groups. It would seem that the more competent "superiors" were remembered by the recipients of the questionnaire and selected for description, while those who were to describe less competent individuals tended to select subordinates.

It is interesting to note that no respondent who was asked to describe a No. 1 person—the best—described a co-worker, and that fewer co-workers were described than either superiors or subordinates. Although there is the same proportion of co-workers as subordinates in levels 6–10, the mean is lower. This is largely a function of the zero frequency at level 1.

The number of descriptions of students is small, but no unusual trends are noted in manner of selection.

Received October 15, 1952.

THE JOURNAL OF APPLIED PSYCHOLOGY
Vol. 37, No. 4, 1953

Rating Patterns for Maximizing Competition and Minimizing Number of Comparative Judgments Necessary for Each Rater

Ray H. Simpson

Department of Educational Psychology, University of Illinois

The problem considered in this paper is represented by a situation which contains the following features:

a. A large number of individuals or written products are to be ranked.

b. The time required to evaluate and give a rank to each individual or product is of such magnitude that any single judge cannot be asked to rank more than 3, 4, or 5 individuals or products.

c. The researcher or administrator wants to use a large number of judges in order to maximize the reliability of the combined rating or ranking.

d. Each individual or product should directly compete with as many other individuals or products as possible.

e. The total group of individuals or products with which a particular individual or product competes should represent a random sample of the total group of competitors against whom he is to be ranked.

The method of paired comparisons would obviously not be appropriate since it would take too much time and be too fatiguing to the judges. Forty individuals or products to be ranked, for example, would involve $n(n-1)/2$ or 780 pairs to be compared by each judge.

Consideration was given to the possibility of using a technique suggested by Guilford[1] which involves selecting from all the individuals or products a limited number to become the basis for a scale. This alternative was not used because of the difficulty in selecting appropriate products or individuals at approximately equal intervals along the scale and because of the volume of work still required of each judge.

The technique described by Uhrbrock and Richardson[2] was also considered. In it one breaks the individuals to be rated into four groups. Within each group each man is compared with every other man in his group and also with each of five "key" men. These latter five form a type of interlocking yardstick. While this method is much more economical of the time of judges than the paired comparison technique, it still was too time consuming to meet all of the requirements of the writer's situation. As a result, the patterns described below have been developed.

Rating Patterns

Table 1 illustrates a pattern to be used when: (a) there are 21 individuals or products to be ranked from highest to lowest; (b) there are to be 21 raters (A through U); (c) each individual or product is to compete with

Table 1

Interlocking Design for Having 21 Raters Give Competitive Ranks to 21 Individuals or Products

Rater	Individuals or Products to be Rated				
A	1	2	7	9	19
B	2	3	8	10	20
C	3	4	9	11	21
D	4	5	10	12	1
E	5	6	11	13	2
F	6	7	12	14	3
G	7	8	13	15	4
H	8	9	14	16	5
I	9	10	15	17	6
J	10	11	16	18	7
K	11	12	17	19	8
L	12	13	18	20	9
M	13	14	19	21	10
N	14	15	20	1	11
O	15	16	21	2	12
P	16	17	1	3	13
Q	17	18	2	4	14
R	18	19	3	5	15
S	19	20	4	6	16
T	20	21	5	7	17
U	21	1	6	8	18

[1] Guilford, J. P. *Psychometric methods.* New York: McGraw-Hill, 1936.

[2] Uhrbrock, R. S. and Richardson, M. W. Item analysis. *Personnel J.*, 1933, **12**, 141–154.

Table 2

Abbreviated Rating Patterns with Each Judge Rating Five Individuals or Products

Total Number to be Rated—25

Judge	Rates				
A	1	2	4	8	16
B	2	3	5	9	17
C	3	4	6	10	18
.					.
.					.
.					.
W	23	24	1	5	13
X	24	25	2	6	14
Y	25	1	3	7	15

Total Number to be Rated—26

Judge	Rates				
A	1	2	4	8	13
B	2	3	5	9	14
C	3	4	6	10	15
.					.
.					.
.					.
X	24	25	1	5	10
Y	25	26	2	6	11
Z	26	1	3	7	12

Total Number to be Rated—27

Judge	Rates				
A	1	2	4	8	17
B	2	3	5	9	18
C	3	4	6	10	19
.					.
.					.
.					.
Y	25	26	1	5	14
Z	26	27	2	6	15
AA	27	1	3	7	16

Total Number to be Rated—28

Judge	Rates				
A	1	2	4	8	13
B	2	3	5	9	14
C	3	4	6	10	15
.					.
.					.
.					.
Z	26	27	1	5	10
AA	27	28	2	6	11
BB	28	1	3	7	12

Total Number to be Rated—29

Judge	Rates				
A	1	2	4	8	18
B	2	3	5	9	19
C	3	4	6	10	20
.					.
.					.
.					.
AA	27	28	1	5	15
BB	28	29	2	6	16
CC	29	1	3	7	17

Total Number to be Rated—30

Judge	Rates				
A	1	2	4	8	22
B	2	3	5	9	23
C	3	4	6	10	24
.					.
.					.
.					.
BB	28	29	1	5	19
CC	29	30	2	6	20
DD	30	1	3	7	21

Total Number to be Rated—31

Judge	Rates				
A	1	2	4	8	16
B	2	3	5	9	17
C	3	4	6	10	18
.					.
.					.
.					.
CC	29	30	1	5	13
DD	30	31	2	6	14
EE	31	1	3	7	15

Total Number to be Rated—32[1]

Judge	Rates				
A	1	2	4	8	16
B	2	3	5	9	17
C	3	4	6	10	18
.					.
.					.
.					.
DD	30	31	1	5	13
EE	31	32	2	6	14
FF	32	1	3	7	15

[1] Similar designs with the same initial sequence can be used with groups larger than 32.

as many other individuals or products as possible; and (d) no judge or rater is to be asked to rank more than five individuals or products. The time involved in the job of rating, reading, or observing necessitates this last requirement.

Examination of Table 1 will show that individual or product number 9, for example, is in competition with each of the 20 other numbers. This means that the final rank for number 9 is as fair a rating as one can get without excessive work on the part of individual judges.

This particular design was used by the writer in a class of 21 students where each student studied and rated, on the basis of criteria previously established, the learning products of five peers. For each student to rank more would have been too fatiguing. The sum of all ranks assigned a particular individual was used to rank the 21 students in the class. Other patterns, based on similar designs, are represented in essence in Table 2.

The technique and patterns suggested here have wide applicability. A few of the additional areas where use would be feasible would be the following: ranking personnel such as in an industrial, military or educational situation; rating short stories; and ranking art forms for esthetic qualities.

Received September 2, 1952.

THE JOURNAL OF APPLIED PSYCHOLOGY
Vol. 37, No. 4, 1953

Some Primary Ratable Characteristics of Instructional Films [1]

Philip Ash [2] **and Thelma R. Hobaugh**

The Pennsylvania State College

A central problem facing the instructional film producer on the one hand, and the teacher on the other, is the evaluation of the film as to its adequacy as a teaching aid.

The *primary* objective of an instructional film is, generally, to instruct. Therefore, the first criterion of the adequacy of a film as an instructional device must be: how well does the film communicate its content—how much is learned from it?

However, a second important objective of an instructional film is to interest, to hold attention, possibly to entertain in the broad sense. In some of its uses, for example as an overview to a unit of instruction, the teacher might intend that the film motivate the students *to want to learn* about the content, with little emphasis on information directly acquired from the film. Furthermore, it seems likely that the extent to which a film is "interesting" in the sense that it gains and holds the attention of the students will influence the extent to which it communicates effectively.

Three kinds of estimates of the quality of an instructional film therefore suggest themselves: first, measures of learning based on test results; second, estimates (judgments, ratings) by people who have seen the film as to the extent to which people will learn from the film; and third, ratings of the affective impact of the film more or less independently of its effectiveness in imparting factual information.

It is the purpose of the present research to determine the extent to which these kinds of estimates are intercorrelated, to answer the question as to the extent to which, from a *rating* of teaching effectiveness or affective quality, it is possible to predict *measured learning*.

A rough check on the teaching effectiveness of an instructional film may be made by administering, both before and after the film, a test generally covering the content of the film. Gains in test scores would reflect the over-all contribution of the film to the students' fund of knowledge. Such a test, however, would not clearly indicate which *parts* of the film were effective and which were ineffective.

A better method of evaluation might involve the development of a *profile measurement* which shows which *parts* of the film are effectively imparting information and which are less effective or ineffective in this respect. Furthermore, technics are available for obtaining rating profiles of the "attention value" and other affective attributes of a film, and for obtaining judgments of teaching value. Profile methods of summating audience ratings with respect to successive sections of programs (entertainment films, radio shows) have been developed extensively during the past ten years, primarily in the radio and motion picture industries.[3] Little comparable work has been done with instructional films, however.

Film Profile Analysis

Under the auspices of the Instructional Film Research Program at The Pennsylvania State College, Dr. Loran C. Twyford conducted an extensive study of the feasibility of a variety of methods for developing profiles of audience evaluations of the instructional

[1] The research on which this article is based was conducted under the auspices of the Instructional Film Research Program under Contract N6-ONR-269, Task Order VII with the Special Devices Center of the Office of Naval Research. This research is reported in Technical Report SDC 269-7-23, *Film Profiles*. The authors gratefully acknowledge permission to analyze the correlation matrix that evolved out of the Ph.D. dissertation completed by Loran C. Twyford: *A Comparison of Methods for Measuring Profiles of Learning from Instructional Films*, The Pennsylvania State College, 1951; Publication No. 3314, University Microfilms, $2.76. Ann Arbor, Michigan. Dr. Ash was responsible for planning the study and writing the report; Miss Hobaugh was responsible for carrying out the factor analysis.

[2] Dr. Ash is now a member of the Industrial Relations Research Department, Inland Steel Company, East Chicago, Indiana.

[3] An extensive review of profile analysis technics in radio and motion pictures is included in Dr. Twyford's study, *op. cit.*

film. These profiles were graphs, drawn against time as the base-line, of the summed responses of the audience to several dimensions of the film.

The profiles he collected may be classified into three groups:

(1) *Profiles of measured learning and learning gains.* These profiles were based upon an exhaustive test of the film content. The test items were true-false form; every fact presented in the film was covered by an item, and each item could be related back to a specific point (in time, sequence, and footage) in the film.

(2) *Profiles of estimated teaching effectiveness.* For these profiles the subjects were asked: (a) to rate whether they thought they were learning; (b) to rate whether they thought students would learn; or (c) to rate (on a third showing) whether they had learned the content presented. The ratings were made during the film showing, on a five-point scale. The subjects made their responses by means of the Film Analyzer [4] system, which provides for the collection by polygraphic recording of the responses of each of a group of 40 subjects.

(3) *Profiles of affect dimensions.* Profiles were collected on the dimensions: (a) "Like-Dislike the film"; (b) "The film is Clear-Unclear"; and (c) "The film is Good-Bad." These profiles also were made on five-point scales and the judgments were recorded by means of the Film Analyzer system concurrently with the viewing of the film.

Procedures

In all, Dr. Twyford collected data on 276 high-school and college students, divided into seven equated groups.

The students in the experiment met as intact classes. A given class was then randomized into seven groups. Each of the seven groups in a class followed a different one of seven experimental procedures.

The film shown to the experimental population was a 10-minute section of a longer film on precision measuring instruments. The original film had been designed by the Instructional Film Research Program. It is essentially an informational film which undertakes to show various kinds of precision measuring instruments and their method of use. The 10-minute section used in the experiment was a complete and intact section of the original.

Groups I, II, and III each saw the film three times, and made a continuous rating on each occasion. Group IV took Form A of the test after seeing the film once, Group V took Form B after seeing the film once. Groups VI and VII took Forms A and B respectively without seeing the films.

The experimental directions were (in part) as follows:

1. *"I am learning."* Specific instructions given to students making this rating were:

"You are to rate the film on the basis of your learning from the film. When facts are presented by the film that you already know, you should report that you are not learning. Some facts in the film may be so difficult that you could not answer questions about them if you were asked to do so after the film showing. These facts were not learned. Learning means that some change has occurred to you that we could discover by comparing your knowledge before and after the film showing."

2. *"I predict learning."* Specific instructions given to students making this rating were:

"You are to predict the amount of learning you would expect a group composed of your classmates to learn from a single showing of the film. The group will not be learning much if the material being presented is previously known. The group will not be learning if the material being presented is too difficult. Learning may occur when the film presents facts that are new. Learning means that some change has occurred to the person that we could discover by comparing his knowledge before and after the film showing by means of tests."

3. *"I like-dislike."* Students making this rating were given the following instructions:

"You are to rate the film according to the degree to which you like the parts of the film. If you like a part very well, depress key No. 5. If you dislike a part of the film, depress key No. 1. The keys that lie between should be pressed to show just how much you like or dislike each part. It will be possible for us to improve the film by making the entire film similar to the parts that you indicate you like. Rate the film from beginning to end."

During a second showing of the film to the same groups, each group was asked to respond a

[4] At each of the 40 stations (chairs) of the Film Analyzer system a box containing five piano-like keys is located. Each subject presses a key signifying his response, and a simultaneous record of the 40-response channels is made on a moving paper. The five choices are coded so that each response channel produces dashed, solid, or a pair of dashed and solid lines identifying the key being pressed. For a description, see Carpenter, C. R., et al. *The Film Analyzer.* Special Devices Center Technical Report 269-7-15, Instructional Film Research Program, The Pennsylvania State College, 1950.

second time, using the same judgment that had been used the first time.

During the third showing, the students in one group were asked to rate how much they had learned; in the second group they were asked to rate how clear the film was; in the third group they were asked to rate the excellence of the film as a teaching aid.

The remaining four groups of each class were handled as follows:

One group took form A of a true-false test which exhaustively sampled the information contained in the film (in both pictures and commentary).

A second group took form B of the same test (true questions of form A were false questions in form B; false questions in form A were true questions in form B).

These two groups acted as a control on the pre-film knowledge of the experimental population; they did not see the film.

The two remaining groups received no pre-tests. They were each shown the film, and then were respectively given one of the two test forms mentioned above.

The nine rating profiles (Groups I–III) were each established by taking the average rating for the group for every two feet of film, a total of 151 points for each profile.

The test questions were distributed approximately evenly to provide a measure in each interval. The four basic test profiles (for Groups IV–VII) were prepared by calculating the mean score for each point on the time axis. Some questions were used at more than one point in the film. These four basic profiles generated a variety of profiles of learning gains (difference between a post-test profile and an appropriate pre-test profile).

The list of the profiles is as follows:

1. Footage—number of feet of elapsed film, measure of the time base.
2. I am learning—1st showing: Ratings of extent of learning during first viewing of film by Group I.
3. Like-Dislike—1st showing: Ratings of degree to which viewer liked the film, during first viewing of film by Group II.
4. Predicted learning—1st showing: Ratings of the extent to which one's classmates would learn the content, during the first viewing of the film by Group III.
5. I am learning—2nd showing: As for profile *2*, during second viewing by Group I.
6. Like-Dislike—2nd showing: As for profile *3*, during second viewing by Group II.
7. Predicted learning—2nd showing: As for profile *4*, during second viewing by Group III.
8. Rated end knowledge—3rd showing: Ratings of amount learned during previous showings, during third viewing by Group I.
9. Clarity: Ratings of the extent to which the film presentation was clear and understandable, during third viewing by Group II.
10. Good-Bad film: Ratings of the excellence of the film as a teaching aid, during third viewing by Group III.
11. Test A learning: Mean score of Group IV minus mean score of Group VI (control or no-film group for Form A) for each profile point.
12. Test B learning: Mean score of Group V minus mean score of Group VII (control or no-film group for Form B) for each profile point.
13. True questions learning: Mean score on true questions from Groups IV and V minus mean score on true questions from Groups VI and VII.
14. False questions learning: As for profile *13*, but for false questions.
15. Comparable questions learning I: Mean score for Group V minus mean score for Group VI, for comparable questions.
16. Comparable questions learning II: As for profile *15*, Group IV minus Group VII.
17. Maximum learning: Profile obtained by assuming maximum learning.
18. Terminal knowledge, Form A: Group IV mean scores for profile points.
19. Terminal knowledge, Form B: As for profile *18*, for Group V scores.
20. Terminal knowledge, true questions: Group IV mean scores and Group V mean scores on true questions only.
21. Terminal knowledge, false questions: As for profile *20*, for false questions.

Purpose of Present Analysis

Dr. Twyford was primarily interested in developing profile construction methodology, and in comparing the relative efficiency, with *measured learning* as a criterion, of the *affect rating* profiles and the *estimated teaching effectiveness* profiles for detecting the strong and weak points of the film. Part of his analysis, therefore, involved correlating the profiles. This was done by taking, for each profile, a measurement at every two feet of film length (151 measurements in all) and intercorrelating the profile arrays.[5]

[5] Before further analysis, Dr. Twyford applied a series of corrections to his correlations. These corrections were not employed in this analysis. These corrections for "drift," "lag," and "carry-over" involved either smoothing lines or partialling out time. They were not applied in this factor analysis; first, because they had little effect on the observed correlations; second, because the partial correlations would be taken care of in the factor matrix; and

Table 1

Intercorrelation Matrix for Twenty Profiles and Film Footage

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21
1. Footage		33	41	49	30	29	50	06	02	43	−05	−11	−07	−03	−04	−01	−13	−23	−21	−22	−23
2. I am Learning—1st Showing	33		65	89	90	48	91	37	55	89	33	20	27	26	22	30	12	02	09	05	04
3. Like-Dislike—1st Showing	41	65		76	59	67	71	58	62	73	03	02	−01	05	04	07	−02	−06	−02	−01	−09
4. Predicted Learning—1st Showing	49	89	76		76	55	91	39	53	89	17	07	14	12	10	17	−01	−11	−06	−07	−12
5. I am Learning—2nd Showing	30	90	59	76		47	86	36	52	85	35	24	25	32	27	35	21	07	14	10	11
6. Like-Dislike—2nd Showing	29	48	67	55	47		58	55	50	58	11	03	07	13	09	16	−01	−02	02	02	−07
7. Predicted Learning—2nd Showing	50	91	71	91	86	58		37	53	94	27	12	19	23	16	28	06	−07	−01	−02	−06
8. Rated End Knowledge—3rd Showing	06	37	58	39	36	55	37		78	50	15	15	12	16	08	15	04	09	12	10	05
9. Clarity	02	55	62	53	52	50	53	78		69	16	15	12	15	12	17	07	11	14	16	05
10. Good-Bad Film—3rd Showing	43	89	73	89	85	58	94	50	69		25	18	20	24	20	29	08	−01	05	03	00
11. Test A Learning	−05	33	03	17	35	11	27	15	16	25		74	82	90	71	78	71	60	59	55	65
12. Test B Learning	−11	20	02	07	24	03	12	15	15	18	74		85	84	77	66	75	62	69	56	72
13. True Questions Learning	−07	27	01	14	25	07	19	12	12	20	82	85		72	70	67	68	54	55	49	59
14. False Questions Learning	−03	26	05	12	32	13	23	16	15	24	90	84	72		76	75	77	66	68	62	72
15. Comparable Questions Learning I	−04	22	04	10	27	09	16	08	12	20	71	77	70	76		72	75	59	52	50	61
16. Comparable Questions Learning II	−01	30	07	17	35	16	28	15	17	29	78	66	67	75	72		71	50	62	57	52
17. Maximum Learning	−13	12	−02	−01	21	−01	06	04	07	08	71	75	68	77	75	71		84	85	80	84
18. Terminal Knowledge, Form A	−23	02	−06	−11	07	−02	−07	09	11	−01	60	62	54	66	59	50	84		85	92	88
19. Terminal Knowledge, Form B	−21	09	−02	−06	14	02	−01	12	14	05	59	69	55	68	52	62	85	85		87	86
20. Terminal Knowledge, True Questions	−22	05	−01	−07	10	02	−02	10	16	03	55	56	49	62	50	57	80	92	87		77
21. Terminal Knowledge, False Questions	−23	04	−09	−12	11	−07	−06	05	05	00	65	72	59	72	61	52	84	88	86	77	

Table 2

Factor Loadings Before Rotation (F_0)

Variable	I	II	III	IV	V	VI	VII	h^2
1	.156	−.439	−.241	.140	−.334	.153	.094	.438
2	.682	−.573	−.235	.199	.205	−.058	−.077	.939
3	.504	−.668	−.204	−.045	−.188	.114	.029	.793
4	.582	−.700	−.189	.157	−.040	.094	−.123	.915
5	.694	−.493	−.192	.188	.165	−.109	.082	.843
6	.456	−.511	.180	−.192	−.231	−.060	.055	.599
7	.653	−.666	−.265	.188	−.036	−.075	−.012	.983
8	.463	−.397	.426	−.434	.134	−.044	.020	.762
9	.538	−.470	.376	−.258	.266	−.058	−.064	.796
10	.696	−.659	−.119	.094	.092	.026	.034	.952
11	.756	.419	−.272	−.163	.033	−.168	−.119	.891
12	.708	.501	−.165	−.188	.133	.221	.100	.892
13	.680	.429	−.274	−.200	.137	.148	−.173	.832
14	.774	.463	−.180	−.133	−.038	−.100	.099	.885
15	.672	.429	−.213	−.155	−.125	.131	.050	.741
16	.716	.352	−.202	−.107	−.088	−.221	.077	.752
17	.697	.608	.069	.143	−.086	.054	.079	.897
18	.599	.620	.360	.233	−.025	.044	−.135	.948
19	.638	.576	.308	.205	.070	−.078	.136	.905
20	.599	.557	.377	.241	−.079	−.147	−.140	.918
21	.600	.640	.206	.182	.135	.128	.104	.890

Table 3

Rotated Factor Matrix ($V = F\Lambda$)

Variable	Test Scores Factor (II′)	Rating Scales Factor (III′)	Terminal Knowledge Factor a (V′)	Terminal Knowledge Factor b (VII′)	(I′)	(IV′)	(VI′)
1	−.070	**.344**	−.093	.006	.062	−.231	.223
2	.106	**.808**	**.436**	−.095	−.080	−.175	−.051
3	−.010	**.746**	−.026	.068	.001	−.019	.177
4	.010	**.835**	.234	.031	−.140	.190	.133
5	.175	**.759**	**.414**	−.037	.078	−.158	−.103
6	.057	**.678**	−.111	.392	.036	.151	−.019
7	.085	**.833**	.268	−.013	−.033	−.214	−.031
8	.027	**.685**	−.022	**.423**	.029	.469	−.064
9	.020	**.788**	.199	.340	−.047	.313	−.094
10	.084	**.899**	.314	.059	.026	−.090	.044
11	**.787**	.055	.088	−.022	−.121	.259	−.164
12	**.780**	−.004	.133	.005	.111	.309	.197
13	**.709**	.013	.101	−.095	−.166	.305	.136
14	**.850**	.044	.087	.094	.097	.231	−.098
15	**.773**	−.005	−.018	.079	.043	.224	.146
16	**.753**	.087	.054	.083	.069	.180	−.206
17	**.928**	−.063	.273	.300	.090	−.024	.029
18	**.864**	−.057	.391	**.484**	−.108	−.095	−.012
19	**.838**	−.003	**.433**	**.406**	.163	−.057	−.138
20	**.839**	−.010	.367	**.524**	−.119	−.116	−.192
21	**.830**	−.091	**.427**	.280	.134	−.028	.063

Table 4

Cumulative Transformation Matrix (Λ)

Variable	I′	II′	III′	IV′	V′	VI′	VII′
I	.001	.735	.579	.089	.303	−.003	.290
II	.018	.625	−.777	.102	.000	.039	.000
III	.034	.022	.223	.056	.162	−.072	.828
IV	.012	.096	.000	−.978	.718	−.028	.000
V	.058	−.244	.102	.150	.605	−.126	−.480
VI	.011	.000	.000	.000	.000	.988	.000
VII	.997	.000	.000	.000	.000	000	.000

Table 5

Cosine Matrix ($C = \Lambda\Lambda'$)

Variable	I′	II′	III′	IV′	V′	VI′	VII′
I′	1.000	.000	.000	.000	.050	.000	.000
II′	.000	1.000	−.080	.003	.148	.000	.348
III′	.000	−.080	1.000	.000	.273	.000	.300
IV′	.000	.003	.000	1.000	−.576	.000	.000
V′	.050	.148	.273	−.576	1.000	−.108	−.068
VI′	.000	.000	.000	.000	−.108	1.000	.000
VII′	.000	.348	.304	.000	−.068	.000	1.000

It is the purpose of the present study to analyze the resulting correlation matrix, to determine whether the domain as tapped by the twenty profiles can be readily reduced to a fewer number of dimensions. It is obvious that the measured learning profiles at least must intercorrelate fairly highly, since they were all based on the same instruments. It was the initial hypothesis of this analysis, therefore, that the whole matrix could be reduced to three factors: a measurement factor, with loadings on the tests; a learning factor, with loadings on both the tests and the estimated teaching effectiveness ratings; and an affective rating factor.

Results

Table 1 presents the intercorrelation matrix. A centroid analysis was applied, yielding the (unrotated) solution given in Table 2. Rotation yielded the final factor matrix in Table 3. The transformation and cosine matrices are given in Tables 4 and 5.

Very early in the rotations, two factors practically orthogonal to each other with large loadings emerged. One included all the test profiles (II′), the other included all the rating profiles (III′). For neither of these, nor for any other factor, was the projection of film footage (a measure of the time base) significant, indicating that these were not simply functions of sequence.

Repeated rotational efforts failed to lead to the identification of any other significant factors. There is a suggestion of overlap between ratings of learning and measured learning in factors V′ and VII′, but the loadings are too small to warrant any large degree of confidence.

Summary

This analysis suggests that test-measured learning gains from a film are more or less independent of ratings of learning or ratings of affective qualities of the film. Furthermore, the results indicate that all the rating scales used have a large common-factor variance. Whether viewers rate on an "I am learning" scale, a "Like-Dislike" scale, a

third, because the corrections were applied not to the entire correlation matrix, but to different sectors of it.

"Good-Bad Film" scale, or any similar scale, much the same function seems to be rated.

The findings of this analysis do not agree completely with the conclusion of the original study that a combined test measure is predicted to a significant extent by ratings on the scale "I am learning." Examination of the correlation matrix before and after corrections for "lag," "drift," and "carry-over" (see footnote 5) shows that correlations between this rating scale and the test score profile were uniformly stepped up as a result of the corrections. It is possible that inclusion of these corrections would have led more definitely to the establishment of a common learning factor, with saturations both on the tests and on the scales rating learning functions.

Received October 27, 1952.

THE JOURNAL OF APPLIED PSYCHOLOGY
Vol. 37, No. 4, 1953

The Relation of Light Intensity to Accuracy of Depth Perception

A. S. Edwards

The University of Georgia

The following experiments were performed to discover the relation of light intensity to the accuracy of visual depth perception.

Procedure

In the first three experiments two lights were used, one of 50 fc, the other of 12 fc. In the fourth experiment, three intensities of light were used, 12 fc, 50 fc, and 100 fc. With all lights, shadows were eliminated.

Apparatus. The apparatus was a modified Howard-Dolman depth perception apparatus. Instead of the upright poles, two blocks of wood were substituted. These blocks were made to hold rectangles of heavy pasteboard on which were pasted gray or colored paper. The size of the stimulus thus presented was 1½" by 2¼". The headrest was at 20 feet distance and kept the *S* from seeing the tops of the blocks. A standard gray (Stoelting gray No. 19) was always used. The movable stimulus had one of the Stoelting grays or colors as follows: gray No. 19, blue No. 13, green No. 8, yellow No. 4, or red No. 1. Intensities of lights were checked by means of a new light meter from the Physics Department and also by means of one from the local electric power company. New papers were purchased for the experiments.

Subjects. The *Ss* were students chosen at random from classes in the University. Only those with at least 20/20 vision, or vision corrected to 20/20, were used. The Ortho-Rater [1] was used to check vision. Color blind *Ss* were eliminated.

Instructions. *Ss* were told, "The object of the experiment is to see how accurately you can judge distance by aligning a movable object with a stationary one. You are to take these strings and pull the right hand object so that it is exactly even with the stationary gray. The right hand string pulls the movable object forward and the left hand string pulls it backwards. When you think the movable stimulus is exactly even with the stationary one, drop the strings. You may move the right hand object back and forth as much as you want to. Keep your chin on the chin rest."

Experiments 1, 2, and 3. The first series was run with 40 *Ss* with 50-fc light, and with 40 *Ss* with 12-fc light. In the second series there were 30 *Ss* who were used as both control and experimental *Ss* with the *12-fc light first*. The third series had 30 *Ss* used as both control and experimental *Ss*, but with the *50-fc light used first*.

Experiment 4. This experiment was carried out under the direction of the author by Mr. E. L. Franklin, a graduate student. There were 50 *Ss* and a 100-fc light was used in addition to the 12-fc and 50-fc lights. The order of lights was alternated.

Results

On the average it appears that depth perception is more accurate with the greater amount of illumination. But in the first three experiments, it is only in Experiment 1 that this conclusion is substantiated by a critical ratio that indicated significance at the one per cent level. See Table 1.

In Experiment 4, the averages are better as the light intensity increases. But only the difference between the 12-fc and the 100-fc lights gave a critical ratio that was significant at the one per cent level. See Table 2.

Individual Results. It appears that some *Ss* are more accurate with less rather than with more light. In the first three experiments, the per cent of *Ss* who did better with 12-fc than with 50-fc light were as follows: (a) Experiment 1, 25 per cent; (b) Experi-

[1] Bausch and Lomb Optical Company. *Standard practice in the administration of the Bausch and Lomb occupational vision test with the Ortho-Rater*, 1944, pp. 1–5.

Table 1

Accuracy of Visual Depth Perception with Two Intensities of Illumination

Note: The standard (stationary) stimulus object was at 50 cm. on the scale. M shows the mean on the scale.

		50 fc		12 fc		
Exp.	*Ss*	M cm.	S.D.	M cm.	S.D.	C.R.
1	80	52.47	3.72	54.72	3.1	3.1 (1%)
2	30	53.07	3.76	54.25	2.7	1.42
3	30	55.71	4.35	56.85	3.45	1.09

Table 2

Accuracy of Visual Depth Perception with Three Intensities of Illumination

Note: The standard stimulus was at 50 cm. and the M gives the average on this scale.

		50 fc		12 fc		100 fc	
Exp.	*Ss*	M cm.	S.D.	M cm.	S.D.	M cm.	S.D.
4	50	(1) 56.6	1.92	(2) 56.9	1.4	(3) 55.5	2.88

C.R.: 1–2 = 0.93; 1–3 = 2.29 (5%); 2–3 = 3.18 (1%).

ment 2, 30 per cent; and (c) Experiment 3, 20 per cent.

In Experiment 4, which used three intensities of light, the per cent of judgments that were better with the light of less intensity are as follows: (a) 20 per cent better with 12 fc than with 100 fc; (b) 20 per cent better with 50 fc than with 100 fc; and (c) 22 per cent better with 12 fc than with 50 fc.

In our four experiments, it thus appears that a considerable number of *Ss* judge depth better with less rather than with greater intensities of light.

Summary

1. In terms of averages alone, it appears that *Ss* were more accurate in depth perception with greater illumination rather than with less.

2. Analyses of our data, however, show that from one-fifth to one-fourth of our *Ss* were more accurate with less rather than with more intense illumination.

3. It appeared that no one intensity of illumination was optimal for all *Ss*.

Received November 3, 1952.

THE JOURNAL OF APPLIED PSYCHOLOGY
Vol. 37, No. 4, 1953

Instrument Reading III: Check Reading of Instrument Groups

W. J. White, M. J. Warrick, and W. F. Grether

Psychology Branch, Aero Medical Laboratory, Wright Air Development Center, Wright-Patterson Air Force Base, Ohio

Most previous studies (1, 3) of instrument reading have been concerned solely with precise quantitative readings. However, it has been pointed out by Grether (2) that there are many situations where instruments are read only to obtain assurance of a null, normal or desired indication (check reading), or for an indication of the direction in which an instrument pointer deviates from a desired value or position (qualitative reading). Milton, Jones, and Fitts (4) report that the average duration of the visual fixations of a pilot flying on instruments is approximately 0.5 seconds. This suggests that aircraft instruments are typically read in a check reading manner.

This paper summarizes a series of experiments concerning the effect of variation in pointer alignment position, in dial diameter and in pointer design on check and qualitative reading of instrument groups. For more detailed reports of these experiments, the reader is referred to a USAF Memorandum Report by Warrick and Grether (6) and two USAF Technical Reports by White (7, 8).

Experiment I: Effect of Pointer Alignment Position on Check and Qualitative Reading

A series of studies was carried out to determine the effect of various pointer-alignment positions on the speed and accuracy of check and qualitative reading of a four by four arrangement of 16 instruments as shown in the center panel of Figure 1. In addition to alignment at the 9, 12, 3 and 6 o'clock positions, several mixed alignment positions were studied.

Apparatus and Procedure. The simulated panel of 16 instruments was presented by means of projected lantern slides. By the operation of a shutter device the subject controlled the exposure of the projected instruments. For the check reading tests, the subject responded by operating a three-position toggle switch held in his lap. He was instructed to move the switch to the right if all pointers were aligned, and to the left if any pointers were misaligned. Forty college students were used as subjects in this experiment.

A somewhat more complicated response was required in the measurement of qualitative reading performance. To the right of the subject was a group of 16 three-position toggle switches arranged in the same manner as the 16 simulated instruments. If there was a deviating pointer the subject moved the corresponding switch in a direction to indicate an increase or decrease from the alignment position. Forty aviation cadets and forty experienced instrument pilots were used in the experiment. Each of the eighty subjects served under only one condition of pointer alignment.

The apparatus and procedure used in studying mixed alignment differed from that described above. The stimulus material was an actual mock-up of 16 pointers and dials. Only the 9, 12 and 3 o'clock alignment positions were studied. All pointers in any one row deviated by the same amount from the alignment position, the amount of deviation varying between rows. The exposure shutter was operated by the experimenter, who also set the pointers from the rear of the panel. Twelve subjects, staff members of the Aero Medical Laboratory, were used in this experiment, each subject serving under all experimental conditions.

Results

The results for the check reading study, Table 1, show that the time taken for check reading appears to be relatively independent of the orientation of the pointer. The aver-

Table 1

Speed and Accuracy of Check Reading a Group of 16 Instruments with Homogeneous Pointer Alignment

Note: N = 40 (10 subjects in each group).

Pointer Position	Mean Response Time in Seconds	Per Cent of Responses in Error
9 o'clock	0.66	2.0
12 o'clock	0.69	2.5
3 o'clock	0.68	5.0
6 o'clock	0.63	3.0

Table 2

Speed and Accuracy of Check Reading 16 Instruments with Pointers Aligned Diagonally Around the 9, 12 and 3 o'clock Positions

Note: N = 12.

Pointer Position	Mean Response Time in Seconds	Per Cent of Responses in Error
9 o'clock	1.64	18.8
12 o'clock	1.56	15.6
3 o'clock	1.59	10.6

age times required to scan the panel and operate the lap-held switch were 0.66, 0.69, 0.68 and 0.63 seconds for the pointers in the 9, 12, 3 and 6 o'clock positions respectively. Error frequency likewise did not discriminate between the 9, 12, 3 and 6 o'clock alignment positions. With mixed alignment the response times were more than doubled and the frequency of errors was generally higher as shown in Table 2.

For qualitative reading, the exposure and response time data in Table 3 show that alignment at the 9 o'clock position resulted in more rapid reading of the group of 16 instruments than did alignment at the 12, 3 or 6 o'clock positions. A comparison of the error data shows that the readings at the 9 o'clock alignment position were also more accurate. The results are similar for both pilots and cadets. Mixed alignment conditions caused a considerable increase in response time and errors as shown in Table 4.

Discussion

The apparent superiority of pointer alignment at the 9 o'clock position over the other cardinal positions for qualitative reading may have resulted from the relationship between the direction of the required switch response and the direction of the pointer deviation. The instructions to the subject read in part, ". . . if the reading is too much, flip the switch down, . . . if the reading is too little, flip the switch up. . . ." The direction of the deviation which corresponds to a judgment of "too little" or "too much" changes with alignment position. Apparently the relationship inherent in the 9 o'clock position was optimal under these instructions.

The results of the experiment on check reading did not indicate the superiority of any one of the four alignment positions. They did demonstrate, however, that pointer alignment, particularly uniform pointer alignment, greatly simplifies the task of reading a group of instruments. The response times found in this experiment using 16 instruments are approximately equal to the time required for reading a single instrument in flight.

Experiment II: Effect of Pointer Design on Check Reading

The purpose of this experiment was to determine the effect of certain factors in the design of pointers on the speed and accuracy of check reading a group of 16 instruments with horizontal pointer alignment. The superiority of pointer alignment for rapid check reading apparently results from the simple pattern which is disturbed by any deviating pointer. Any malfunction which would cause the pointer to rotate 180 degrees might not be detected because of the minor resultant change in the general alignment pattern. Variations in the

Table 3

Speed and Accuracy of Qualitative Reading of Sixteen Instruments with Homogeneous Pointer Alignment

Note: N = 48 (12 subjects in each group).

Pointer Position	Mean Exposure Time in Seconds		Mean Response Time in Seconds		Per Cent of Responses in Error	
	Pilots	Cadets	Pilots	Cadets	Pilots	Cadets
9 o'clock	0.90	1.17	1.63	1.84	2.91	5.00
12 o'clock	1.15	1.39	2.02	2.17	6.25	12.08
3 o'clock	1.30	1.83	2.32	2.34	23.33	20.83
6 o'clock	1.62	1.46	1.89	2.29	10.83	9.58

Table 4

Speed and Accuracy of Qualitative Reading 16 Instruments with Pointers Aligned Diagonally Around the 9, 12 and 3 o'clock Positions

Note: N = 12.

Pointer Position	Mean Response Time in Seconds	Per Cent of Responses in Error
9 o'clock	4.51	15.8
12 o'clock	4.36	22.6
3 o'clock	4.78	14.9

design of the base of the experimental pointers in the present study were intended to increase the change in pattern resulting from a 180-degree deviation.

Apparatus and Procedure. The stimulus material and method of presentation and procedures were very similar to those used in the preceding experiments in this report. The five pointer designs studied in this experiment are shown in Figure 1. All pointers had about the same physical dimensions. They were 1⅛" long and 3/32" wide at the center. Five film strips, one for each pointer design, were made of a panel containing 16 simulated 1¾" instruments.

Forty instrument pilots and an equal number of cadets served as subjects; each served under only one condition of pointer design.

Results

The results in terms of speed and accuracy of check reading as a function of pointer design are summarized in Table 5. In the first two columns of this table are the per cent of readings in error, all errors being combined for pilots and cadets. In the next two columns are the per cent of undetected 180-degree deviations. For both subject groups the number of undetected deviations was greatest for design D. For the pilot group only the largest difference, that between designs D and E, is significant at the 5% level of confidence. For the cadet group none of the differences are significant at the 5% level of confidence. This suggests that the experimental pointers (designs A, B, C and E), when rotated 180 degrees from alignment position, were little if any more effective in pro-

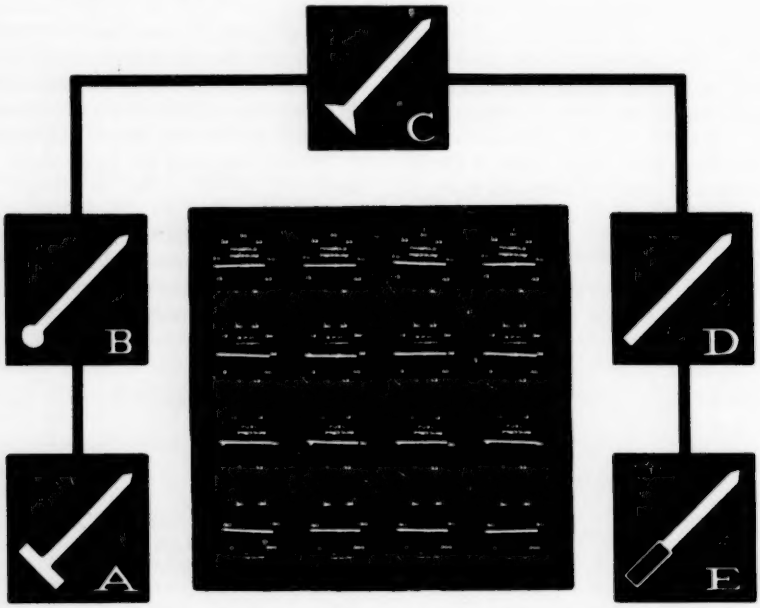


FIG. 1. Instrument panel arrangement used in all check reading experiments and pointer designs used in Experiment II.

Table 5

The Effect of Pointer Design on the Speed and Accuracy of Check Reading 16 Instruments with Homogeneous Pointer Alignment

Note: N = 90 (10 subjects in each sub-group).

Pointer Design	Per Cent of Readings in Error (N = 200 Trials)		Per Cent of Undetected 180° Deviations (N = 40 Trials)		Per Cent Misc. Errors (N = 160 Trials)		Mean Response Time in Seconds	
	Pilots	Cadets	Pilots	Cadets	Pilots	Cadets	Pilots	Cadets
A	14.00	11.00	45.00	37.50	6.25	4.37	2.00	2.11
B	16.50	10.50	47.50	35.00	8.75	4.37	1.62	1.91
C	15.50	15.5	47.50	27.50	7.50	12.50	1.80	2.05
D	14.50	12.5	62.50	47.50	2.50	4.37	1.75	1.90
E	21.50	—	39.00	—	16.00	—	2.80	—

viding cues for detection of misalignment than was the standard pointer (design D).

If we compare the data for miscellaneous errors in Table 5, a somewhat different picture is obtained. For the pilot group the error percentage was lower for the standard pointer, design D, than for other designs with the differences significant at the 5% or greater level of confidence. The results for the cadet group do not support this superiority of the design D. The mean response time values in the last two columns of Table 5 show no significant differences between pointer designs.

Discussion

The failure of the experimental pointers to reduce substantially the 180-degree type error may, perhaps, be attributed to a configurational change brought about when the pointer overlays the dial numerals and graduations. This reduces the visibility of the expansions of the pointer base. This is especially likely when the panel is scanned rapidly as in check reading.

Experiment III: Ocular Movements in Relation to Dial Diameter

The pioneering work of Paterson and Tinker (**5**) on typographical factors which influence reading suggests that there might be an optimal dial diameter for instrument reading purposes. The optimal dial diameter should be such that the resulting instrument panel size could be scanned accurately, rapidly and with a minimum number of eye fixations. The most suitable approach to this problem appeared to be through the combined measurements of manual responses and of eye movements while check reading a panel of instruments.

Apparatus and Procedure. For this experiment three panels were used. The instrument diame-

Table 6

Eye Movements, Response Times and Errors in Check Reading a Group of 16 Instruments

Note: N = 6.

Type of Measure	Dial Diameter		
	1 inch	1¾ inches	2¾ inches
Mean number of fixations per trial	2.8	2.6	2.9
Mean time per fixation	0.29	0.26	0.26
Total scanning time from preliminary to final fixation	0.82	0.72	0.75
Per cent deviating pointer trials on which deviating pointer was fixated	72	71	81
Response time (sec.)	0.76	0.72	0.73
Per cent errors	6	4	12

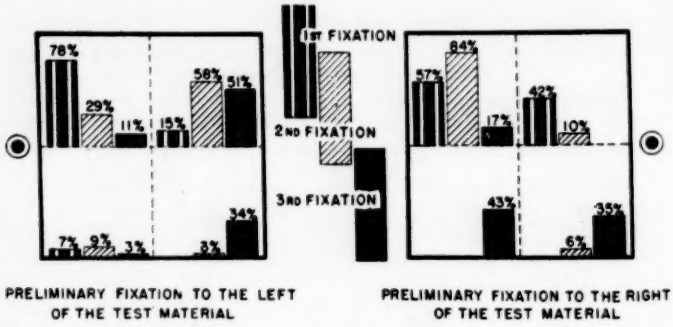


FIG. 2. Distribution of first, second and third eye fixations in check reading a panel of 16 instruments.

ters were 1, 1¾, and 2¾ inches. In all cases the spaces between the margins of the instrument were ⅛ inch. These panels were check read using both eyes. The movements of the right eye were recorded by the corneal reflection technique. An American Optical Company Ophthalmograph was modified so that exposures could be made with either moving or stationary film. The moving film exposures provided data on the number of fixations, and the lateral position and duration of fixations. The stationary film exposures provided additional data on the position of fixation in both the vertical and lateral dimensions, but did not record the duration of fixations.

On each of 60 test trials the subject was required to fixate a spot to the left of a screen covering the panel. When the screen was raised the subject scanned the panel as needed and then fixated on a spot to the right of the screen. He then signalled his judgment as to whether or not a pointer was deviated from the aligned position. For three of the six subjects the initial and final fixation points were reversed. When the film record of the subject's eye movements was projected against a replica of the corresponding instrument panel, the initial and final fixation points provided references for the analysis of the film records.

Results

The major results obtained from the recording of eye movements are summarized in Table 6, which gives check reading time, response time and error data in addition to the various measures of visual performance. Although not statistically significant, the results are quite consistent in showing slightly superior performance with the 1¾-inch dial size.

Analysis of the location of the 1st, 2nd and 3rd fixations, in relation to the panel arrangement, are shown in Figure 2. The major proportion of the first fixations were in the upper left quadrant of the instrument panel. This was true even when the preparatory fixation was at the right of the panel. Surprisingly few fixations were made on the lower two rows of dials, and those that did fall in this area were usually the last fixation in the scanning sequence. Analysis of the errors committed shows that 22 per cent more errors were made when the deviating pointer was in the lower half of the panel. In the fourth row of Table 6 is shown the per cent of deviating pointer trials in which the deviating instrument was fixated.

Conclusions

The following conclusions are suggested by the results of the foregoing experiments:

1. The pointer alignment position (9, 12, 3 or 6 o'clock) has little effect on simple check reading involving the mere detection of a pointer deviation.

2. Uniform horizontal or vertical pointer alignment facilitates instrument check reading.

3. Pointer alignment at the 9 o'clock position is optimum for qualitative reading as defined by the perceptual judgments and manual responses required in these experiments.

4. Pointer modifications of the type employed in the present experiment do not accomplish a satisfactory reduction in difficulty of detecting 180-degree deviation errors.

5. A $1\frac{3}{4}$-inch instrument dial appears superior to either a larger or smaller dial in terms of the number and duration of eye fixations while check reading.

Received August 25, 1952.

References

1. Fitts, P. M. *Engineering psychology and equipment design.* In S. S. Stevents (Ed.), *Handbook of experimental psychology.* New York: John Wiley & Sons, 1951, 1287–1340.
2. Grether, W. F. Instrument reading. I. The design of long-scale indicators for speed and accuracy of quantitative readings. *J. appl. Psychol.*, 1949, **33**, 363–372.
3. Kappauf, W. E. *Studies pertaining to the design of visual displays for aircraft instruments, computers, maps, charts, tables and graphs: A review of the literature.* Engineering Division, Air Materiel Command, Dayton, Ohio, USAF Technical Report No. 5765, April, 1949.
4. Milton, J. L., Jones, R. E. and Fitts, P. M. *Eye fixations of aircraft pilots, II. Frequency, duration and sequence of fixations when flying the USAF instrument low approach system.* Engineering Division, Air Materiel Command, Dayton, Ohio, USAF Technical Report No. 5839, October, 1949.
5. Paterson, D. G. and Tinker, M. A. *How to make type readable.* New York: Harper & Bros., 1940 (out of print).
6. Warrick, M. J. and Grether, W. F. *Effect of pointer alignment on check reading of engine instrument panels.* Engineering Division, Air Materiel Command, Dayton, Ohio, USAF Memorandum Report No. MCREXD-694-17, June, 1948.
7. White, W. J. *Effect of dial diameter on ocular movements and speed and accuracy of check reading groups of simulated engine instruments.* Engineering Division, Air Materiel Command, Dayton, Ohio, USAF Technical Report No. 5826, June, 1949.
8. White, W. J. *Effect of pointer design and pointer alignment on speed and accuracy of reading groups of simulated engine instruments.* Engineering Division, Air Materiel Command, Dayton, Ohio, USAF Technical Report No. 6014, July, 1950.

THE JOURNAL OF APPLIED PSYCHOLOGY
Vol. 37, No. 4, 1953

Dimensional Analysis of Motion: VI. The Component Movements of Assembly Motions [1]

Robert Smader and Karl U. Smith

The University of Wisconsin

The application of electronic methods to investigation of human motions (**1**, **4**, **5**) provides the means of systematic experimental study of long-standing problems of motor coordination. Such electronic techniques, which have been used previously in studying panel-control operations and closely related performances, are used in this study to analyze the component movements in assembly motions. Data are presented bearing upon the relative efficiency of the component movements of assembly motions and upon the effects of practice on such movements.

Two methodological principles are demanded for the scientific investigation of complex human motions. First, the basic dimensions of motion must be subject to control and quantitative specification. Second, the component movements of the pattern must be segregated and measured separately. This second requirement for motion study must be achieved with an economy sufficient to permit replication of the observation of the same motion pattern throughout the course of an experiment.

Control and specification of the various dimensions of movement, especially the bodily and space dimensions of movement, are secured through preplanning of performance situations in order to permit quantitative variation in the direction, plane, magnitude, bodily orientation, and other aspects of motion. Measurement of the component movements in assembly motions may be achieved by means of electronic motion analysis techniques, to be described, which provide, for the first time, methods adequate for the experimental analysis of the separate movements of skilled performances.

[1] This research has been supported by funds voted by the Legislature of the State of Wisconsin, and assigned by the Graduate School Research Committee, the University of Wisconsin.

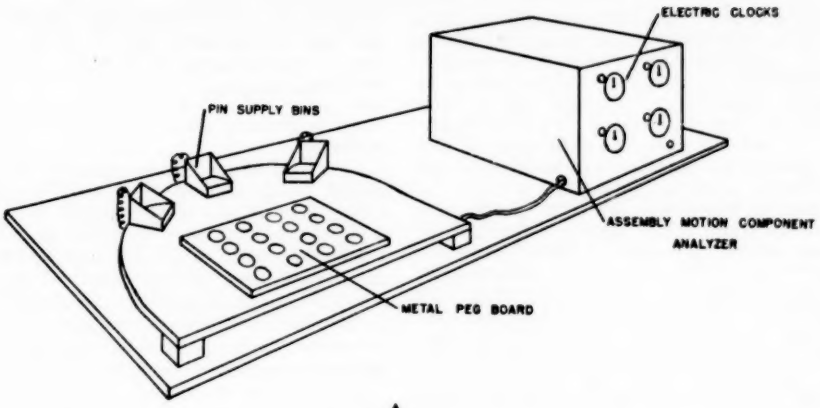
Methods

Figure 1A illustrates the main parts of the apparatus used for electronic registration of the duration of the component movements in assembly motions. The task performed by the subject in this situation, as illustrated by Figure 1B, consists of four major movement components: (a) an initial reaching or travel movement that carries the hand and arm to the bins containing pins used as assembly objects; (b) a grasping movement in securing one of the pins; (c) a return or loaded travel movement that carries the pin to the assembly plate; and (d) a positioning motion of inserting the pin in the assembly plate.

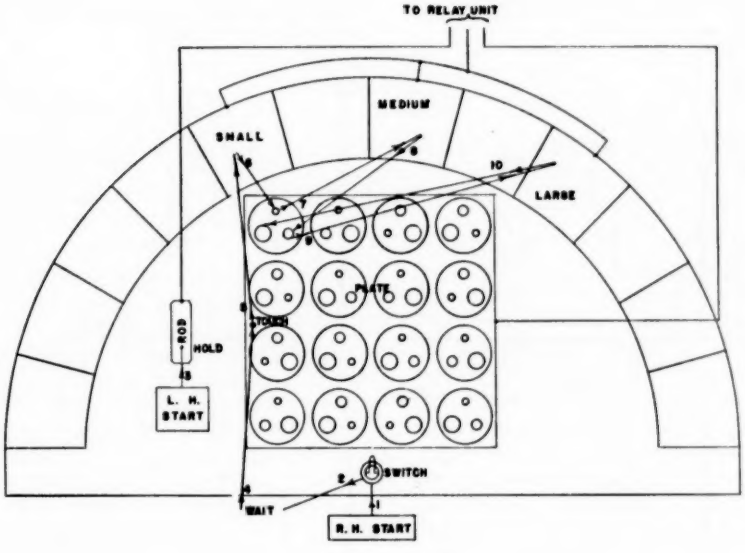
The preplanned performance situation used here is designed to provide for systematic variation in different space dimensions of component movements of the assembly task (Figure 1B). Assembly bins and the assembly plate are arranged generally on a work table measuring 35 in. by 20 in. The metallic assembly bins, which are fitted with metal bases, are 8 in. long, 4 in. wide, and 4 in. high. The floor or base of each bin is curved to facilitate grasping of the pins located inside. Six such bins may be used for pins or collars of different sizes. The distance and arrangement of these bins with respect to the assembly plate may be changed to vary the magnitude, direction and plane of component travel and grasping movements.

The assembly plate (Figure 1B) is of aluminum ½ in. thick, 10½ in. long and 10½ in. wide, drilled with 16 large apertures, $1\frac{15}{16}$ cm. in diameter. Fitted into these apertures are discs which are drilled with three holes of different sizes, i.e., ¼ in., ⅜ in. and ½ in. The discs may be locked into three different positions on the assembly plate. Three sizes of pins, all approximately one inch long but of diameters to fit the three sizes of holes, are provided in three bins. Variations in size of object grasped, serial order of placement of the pins, and other variations in the dimensional aspects of the assembly placing movement are made by means of the different placements of the bins and positions of the holes on the assembly plate.

The subject's movements in this situation are recorded by means of an electronic motion analyzer that automatically segregates and measures separately in 0.01 seconds the duration of each of the four basic component movements in the task. Figure 1A illustrates how the clocks and circuits of the analyzer are arranged with respect



A



B

FIG. 1. A diagrams the general layout of the motion analyzer circuits, the time clocks, the assembly plate, and the bins. The bins are made with curved bottoms in order to facilitate removal of the pins. B shows the surface detail of this same setup. The letters "L.H." and "R.H." in this diagram refer to left hand and right hand respectively. The sequence of movement with the right hand is indicated by number. The size of pins located in the bins is indicated in relation to the position of the bin.

to the work area. The basic principle of operation of the electronic analyzer is that the subject acts as a key in the circuit and thus sequentially activates different relays and clocks [2] during the different stages of the assembly motion. The current level used in the subject's side of the circuit is below cutaneous threshold level, so that the operator has no sensory appreciation of the fact that he is serving as a conductor.

Figure 1B shows how the device performs in actual operation. Upon starting a cycle of assembly motion, the subject initially touches the assembly plate. This contact with the plate starts the first of the four precision time-clocks, the initial travel time-clock. Upon coming into contact with the pins in one of the assembly bins, this first clock stops and the second, the grasping time-clock, starts. When the contact between the pin and the bin is broken, the grasping time-clock ceases to run and the loaded-travel time-clock starts to run. This third clock stops when the pin is brought into contact with the assembly plate, and the fourth clock, the assembly positioning time-clock, starts to run. This fourth clock in turn stops running when electrical contact with the assembly plate is broken and a second cycle of motion is initiated. The different times registered on each clock may be read after one cycle of motion or they may be allowed to summate over several cycles and then be read.

In the experiment to be described, 46 right-handed college students were used as subjects. These subjects were given standard instructions to fill the assembly plate with pins according to a defined sequence, which was kept constant for all subjects. One complete filling of the plate constituted a trial. Subjects performed two trials per day for each of three consecutive days. All scores analyzed were the mean durations of each component movement for each day.

In addition to the running of the test trials, as just described, a calibration trial was run before the first test trial of the second and third days. In these calibration trials, the two manipulation time clocks and the two travel time clocks were electrically connected so that the discrepancy between two clocks in measuring the same elapsed time could be determined.

[2] Model S–1, Standard Electric Time Company, Springfield, Mass.

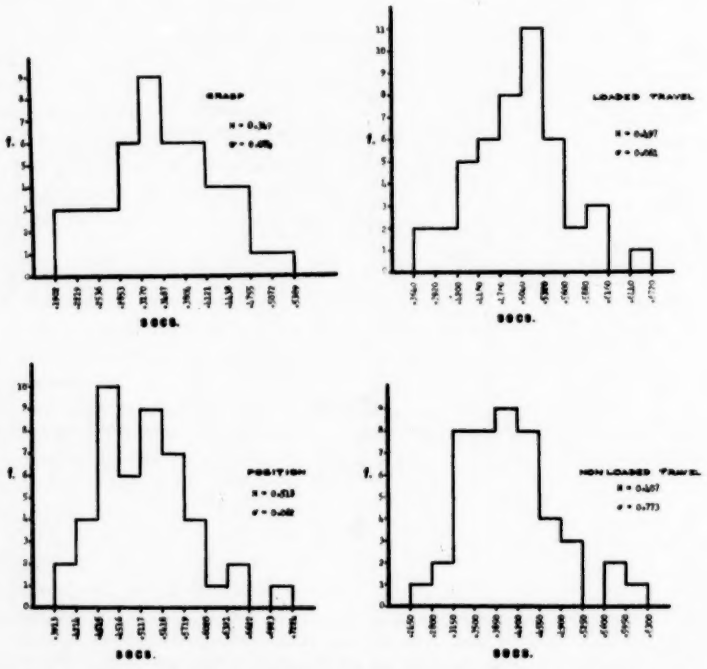


FIG. 2. Distribution of the durations of different component movements in relation to subjects. The figures for the manipulative components are given to the left and those for the travel components to the right.

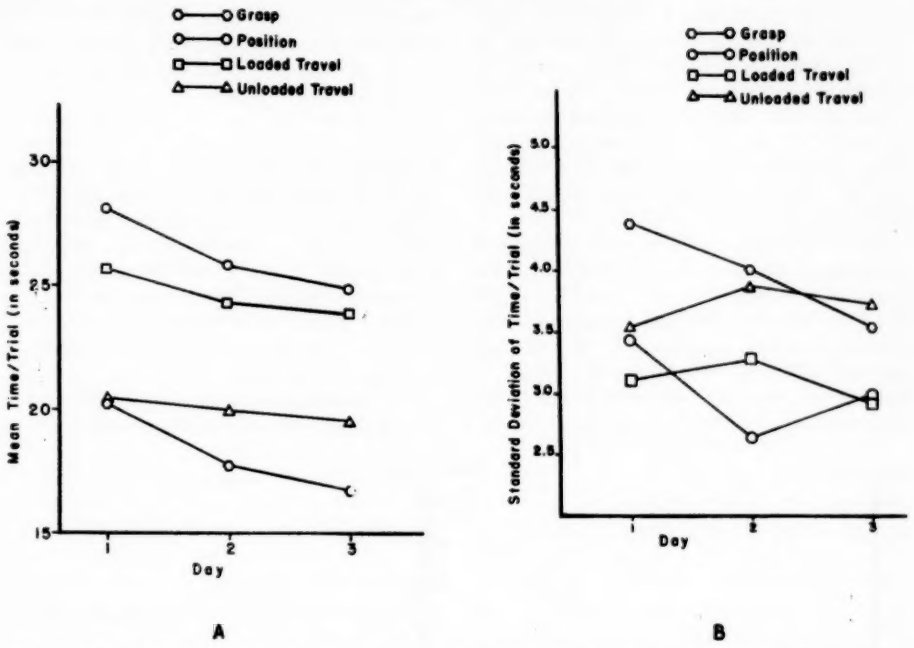


FIG. 3. The change in the duration (A) and in the variability (B) of the four different component movements as a function of practice. Generally, the two travel movements show less change with practice than that observed in the grasping and positioning movements.

Results

The results of this experiment will be presented with respect to the following main points: (1) distribution of time values for different component movements in the assembly task; (2) the effects of practice on different component movements in the task as well as on their interrelation; and (3) the reliability of measurements of different movement components.

Figure 2 shows the distribution of movement times of the four component movements of the assembly task on the third day of practice. The data presented in this figure represent the durations of individual movements for each part of the assembly task. The movement component giving the shortest times is the grasping movement. The assembly positioning movement gives the longest durations. All four movement components show distributions that are slightly skewed positively.

Figures 3A and 3B describe the change in the duration of the component movements of the assembly task as a function of practice. Figure 3A gives the learning curves based on the mean duration of each of the component movements on each day. Figure 3B shows the change in variability of each component movement as a function of practice.

Three of the four movement components in the task show a highly significant change in duration as a result of practice over three days. The greatest change, about 18 per cent, is found for the grasping component of the task. The other manipulative component of the task, the positioning movement, gives an 11 per cent change with practice. Both of these manipulations show a greater change with practice than does either of the travel movements. One of the latter, the unloaded travel movement, shows a statistically significant change with practice only at the 0.05 point. Tests of significance of the differences

between day 1 and day 3, based on "t," indicate that changes occurring in learning for the three other components of movement are significant at the 0.001 point. On each day of practice, the differences between the mean durations of the two manipulative components, grasping and positioning, and between the two travel movements are statistically significant at the 0.001 level.

The variability of different component movements does not change uniformly as a function of practice. The variability of the unloaded travel movement decreases with practice, but not significantly. In contrast, the variability of the grasping motion is reduced sharply with practice. The other manipulative movement, positioning, has its standard deviation changed about 15 per cent with practice. The variability of the loaded travel movement is hardly changed at all during the three days.

The present method of isolating electronically the different components of movement in assembly has provided the means of studying the interrelation between separate movements as a function of practice. In order to conduct such investigations, the individual mean scores for each component of motion are obtained for each day of practice and correlated. Thus, six correlations between pairs of the four component scores are obtained for each day of the study.

The correlation values just described are plotted in Figure 4 as a function of days of

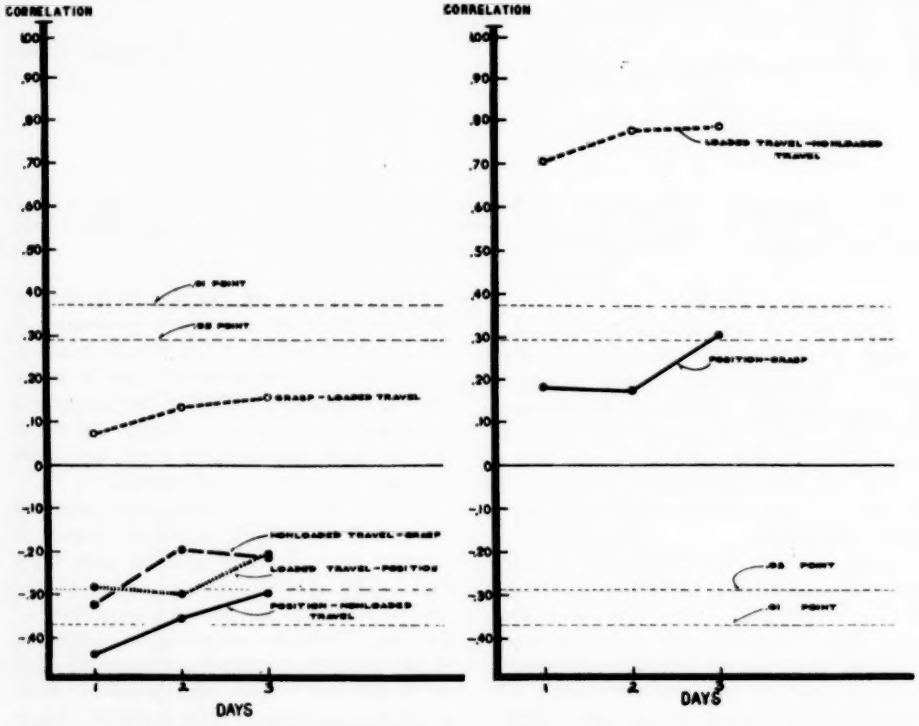


FIG. 4. The level of correlation between different component movements during practice. The plot to the left gives the curves for pairs of movements that are adjacent to one another in the motion pattern. The plot to the right gives the curves for pairs of movements that are non-adjacent in the cycle. The correlation values at the .05 and .01 points are indicated by the dotted lines. Among the different curves, only those for the correlations between loaded and nonloaded travel and between positioning and nonloaded travel are statistically significant for all three days of practice.

Table 1

Coefficients of Stability of the Component Movement Times

Type of Component	Component	Day 1 vs. Day 2 Test-retest Reliability Coefficient	Day 2 vs. Day 3 Test-retest Reliability Coefficient
Manipulative	Position	+0.79	+0.86
	Grasp	+0.79	+0.91
Travel	Unloaded	+0.83	+0.88
	Loaded	+0.80	+0.91

practice. In addition, the broken lines on Figure 4 indicate the 5% and 1% levels at which these correlations may be considered to differ significantly from zero. The correlation functions which give values consistently significant at the 5% level are those representing the relations between loaded and unloaded travel movements and between position and unloaded travel movements. Although it might be expected that the position and grasp movements would correlate highly with one another, these two movements in fact show no significant relation with one another.

The main point to be observed in Figure 4 is the change in the correlation between component movements in relation to practice period. It will be observed that none of the correlations between different pairs of movements is significantly altered during the three days of practice. In other words, the interrelation of movement components does not vary during learning and improvement in performance in the over-all task. If the correlation between component movements is thought of as a measure of integration between these movements, then the results show that learning has little or no effect on the integration of separate movements in assembly motions.

Inasmuch as the present experimental techniques represent quite complex procedures for automatic measurement of behavior, it is of importance to examine the temporal consistency of scores obtained. It will be remembered that each subject performed two trials on each of the three days of practice, and that four scores, representing different component movements, were obtained on each trial. To obtain correlation coefficients between days for each of the four component movements in the task, a mean score for each subject for each type of movement was obtained and these values for different days were then correlated with one another. Table 1 summarizes the values obtained. The table shows that a high degree of consistency is found for the component movement times in the assembly task. It should be noted again that the figures given in Table 1 are based upon only two trials of performance per day and that the figures represent consistency between days. These figures in Table 1 show that the low intercorrelations between component movements, as given in Figure 4, cannot be accounted for in terms of unreliability of the individual movement times.

Summary

This study describes electronic techniques which make possible detailed experimental study of the component movements in assembly motions. Besides indicating the nature and value of these methods for analysis of industrial motions, the present paper describes experimental results bearing upon the characteristics of component movements in an assembly task, and upon the effects of practice on the duration and relation between these movements.

By means of the electronic techniques described, the component movements of travel, grasp, loaded travel and positioning in a unimanual assembly task may be isolated and their durations measured in hundredths of seconds. Distributions of the component movement times for forty-six college student operators are approximately normal.

Practice does not affect uniformly different types of movement in the assembly task. The efficiency of the two manipulative movements, positioning and grasping, are changed the most by practice. In contrast, the travel motions in the task, especially the unloaded travel component, show very little change with practice. Similar results are found for measures of variability of response during practice.

It has been observed also that practice does not alter significantly the correlation between

component movements in the assembly task. Early in learning the correlations between the different component movements of manipulation and travel are near zero. As practice continues, the values of these correlations do not change significantly. At the start of learning, a correlation of about +0.07 is found between the two travel movements in the task. With practice, this correlation increases slightly but not significantly.

The results on the effects of practice, as just described, are equivalent to experimental findings obtained in previous studies of learning and component movements in tracking behavior (2, 3) and panel control motions (1, 4, 5, 6). In all of these tasks, the human body seems to act like a many-channeled system in the production of discrete unrelated movements for a given task, the integration of which is not altered materially by continued practice. The findings described here, along with similar results found for other motion patterns, point toward the lack of significance of learning concepts in understanding the details of movement coordination in skill. The same results, however, point up the great significance of component and dimensional motion analysis in dealing practically and theoretically with problems of skilled motion and the instrumental relations of such motion.

The electronic methods of motion analysis described in this report provide, for the first time, economical methods of obtaining reliable measures of movement components in assembly skills. These methods, along with procedures of preplanning and quantitative control of the dimensions of motion in work, lay the foundation of scientific study of motion in terms of modern experimental designs.

Received October 6, 1952.

References

1. Davis, R., Wehrkamp, R., and Smith, K. U. Dimensional analysis of motion: I. Effects of laterality and movement direction. *J. appl. Psychol.*, 1951, **35**, 363–366.
2. Lincoln, R. S., Simon, J., and DeCrow, T. W. Effects of practice upon different component movements in tracking. *Per. and Mot. Skills Res. Exch.* (In press).
3. Lincoln, R. S., and Smith, K. U. Systematic analysis of factors determining accuracy in visual tracking. *Science*, 1952, **116**, 183–187.
4. Rubin, G., and Smith, K. U. Learning and integration of component movements in a pattern of motion. *J. exper. Psychol.* (In press).
5. Smith, K. U. and Wehrkamp, R. A universal motion analyzer applied to psychomotor performance. *Science*, 1951, **113**, 242–244.
6. Wehrkamp, R. and Smith, K. U. Dimensional analysis of motion: II. Travel distance effects. *J. appl. Psychol.*, 1952, **36**, 201–206.

THE JOURNAL OF APPLIED PSYCHOLOGY
Vol. 37, No. 4, 1953

The F minus K Index on the MMPI

Angus G. MacLean, Arthur T. Tait, and Calvin D. Catterall

California Test Bureau, Los Angeles

Hunt (2) compared F-K raw scores for honest and dissembled MMPI profiles from the same subjects and found that an F-K cutting score of − 11 and below was fairly effective in spotting fake-good cases, but also picked up too many supposedly honest cases. Gough (1) reported the distribution of F-K scores for a group of 691 adult normals, the middle 80% of whom obtained approximately normally distributed scores ranging from − 22 to + 11 with a mean of about − 9.

The present data consist of a sample of 100 candidates for nursing randomly drawn from a large number of similar cases in the Los Angeles and San Francisco Metropolitan areas. These candidates are almost without exception female, are finishing the 12th grade or are in their first year of college; if not attending college, they range from 18 to 22 years of age with a modal age of 18 to 19. These cases provide data worth noting since they are in a selection situation; furthermore, the balance of their MMPI profiles suggests that they are motivated to present themselves in the best light. Few T-scores exceed 60 on any clinical scale and virtually none exceed 70.

The F, K, and L scores for these subjects are distributed as follows: Actual distribution of F scores: T = 50: 64 cases; T = 53: 17 cases; T = 55–59: 8 cases; T = 60–64: 10 cases; and T = 73: 1 case.

A mean and standard deviation should not be calculated for such a distribution.

The distribution of K scores had a mean of 57.9 and a standard deviation of 8.5. The total range was from 30 to 79, and the middle 80% of cases lay between 46 and 69.

The L scale had a mean of 72.0 and a standard deviation of 8.1. The total range was from 50 to 84, and the middle 80% of cases lay between 61 and 82. The scores were slightly negatively skewed, and it seems possible that these candidates, while defensive on the F and K items, felt that frankness on the L items would not be damaging, or even that they might be "catch" items.

The distribution of raw scores, somewhat smoothed, is shown in Table 1.

It is suggested that a cutting score of − 17 might be tentatively adopted: this would include 31% of the nurse candidates but only about 12% of the "Adult Normal" cases discussed by Gough. A cutting score of − 14 would include 50% of nurse candidates and about 22% of Gough's "Adult Normals." It may be assumed, lacking information to the contrary, that the "Adult Normals" consisted merely of non-pathological cases, and included quite a wide range of cases from the moderately self-critical and fairly honest to the highly ego-defensive. About 1% of Gough's cases obtained F-K scores of − 22 or more. Thus, if a cut-off is established at − 17, about 12 or 15% of Gough's cases will be included, it is true, but it may well be supposed that these cases are not so much unusually well-adjusted, as those with a tendency to dissemble even in a non-selection situation.

In view of the above data, the following ranges for the F-K index are suggested:

Table 1

F-K Raw Score Distribution on MMPI Based on a Sample of 100 Candidates for Nursing

Sten Score (1 sten = 0.5σ)	Raw Score (Lower Limits)	% in Category (Smoothed)	Cumulative %
9	−25 up	2.3	2.3
8	−22	4.4	6.7
7	−20	9.2	15.9
6	−17	15.0	30.9
5	−14	19.1	50.0
4	−11	19.1	69.1
3	− 9	15.0	84.1
2	− 7	9.2	93.3
1	− 2	4.4	97.7
0	−1, 0, or positive	2.3	100.0

1. Positive Raw Score, especially if greater than + 2: malingerers or very self-critical and unusually honest subjects.
2. Raw Score from zero to − 10: Normal Area.
3. Raw Score from − 11 to − 16: Doubtful Area.
4. Negative Raw Score greater than − 17: Probable "Fake-good" Area, especially when the F-K difference lies in the twenties.

Optimal cut, as always, refers to a specified pair of populations, and caution must be used in generalizing to dissimilar groups and situations. For example, in another study we found the F-K raw-score difference to range from − 25 to + 30, with a middle-80% range from − 16 to + 9, a mean of − 5 and a median of − 6. The SD for this group is approximately 10 raw-score points. We can only assume that some members of this group desired to appear in a good light, while others were either malingering or unusually self-critical. This would suggest regarding small positive scores as indicative of frankness and insight, while the suspicion of malingering would pertain to positive scores in the neighborhood of + 15 or more.

Received February 11, 1953.
Early publication.

References

1. Gough, H. G. The F minus K dissimulation index for the Minnesota Multiphasic Personality Inventory. *J. consult. Psychol.*, 1950, **14**, 408–413.
2. Hunt, H. F. The effect of deliberate deception on Minnesota Multiphasic Personality Inventory performance. *J. consult. Psychol.*, 1948, **12**, 396–402.

THE JOURNAL OF APPLIED PSYCHOLOGY
Vol. 37, No. 4, 1953

Psychological and Personal History Data Related to Accident Records of Commercial Truck Drivers *

James W. Parker, Jr.

Tufts College

This study is one phase of a larger over-all research project being carried out in the Department of Psychology at North Carolina State College, Raleigh. The general purpose is to improve the method of selecting truck drivers to be employed by a large East Coast trucking concern.

Criterion data for studies such as this have long been a subject of controversy. Up to the present time the data used have been the accident records. The manner in which these data have been used has been varied, but no really "new" idea in criterion data has been devised. It is basically the accident record, therefore, which was used in this study, and the records for the various subjects were equated on the basis of the number of miles driven by each.

The subjects were 104 drivers who were still employed by the company on a certain cut-off date, and who had been trained at the Driver Training School operated by the North Carolina State College. All subjects had been tested at the training school by an examiner trained in psychological testing, and under standardized conditions. In most cases the test results were not used in the employment procedure by the company. Some subjects did not have a complete report of their test performance in their files, and this necessitated using varying sample sizes for the different variables.

The factors or variables with which the accident records are compared are divided into two categories: (a) psychological test data and (b) personal history data which were derived from the application blank filled out by the employee at the time of his employment by the company. The tests used were some of the better-known paper-pencil tests of intelligence, mechanical comprehension, personality, and vocational interest. The results of visual screening with the Bausch and Lomb Ortho-Rater were also included. Personal history data included such things as age, education, marital status, blood pressure, etc.

Two sources were used to arrive at the criterion data: the number of accidents a man had incurred during a certain period of time, and the total number of actual miles driven during this same period of time. The accident rate was then put on the basis of the number of accidents per 5,000 miles. This was done for two types of accidents, those classified preventable and non-preventable by the safety department of the trucking company.

Procedure

The distributions of scores for the entire sample of 104 drivers on the psychological test data and the personal history data were divided into two groups, accident group and non-accident group, with respect to each of the two criteria, preventable and non-preventable accidents per 5,000 miles. The accident group for each of the criteria was further divided into upper and lower halves, excluding the accident-free group. Figures 1 and 2 show the distributions of preventable and non-preventable accidents.

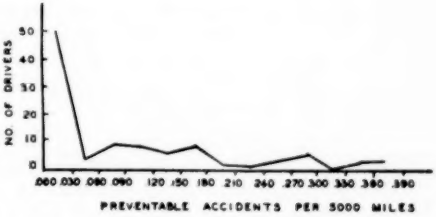


FIG. 1. Distribution of preventable accidents.

* This study is a condensation of a thesis submitted to the North Carolina State College of Agriculture and Engineering of the University of North Carolina, Raleigh, in partial fulfillment of the requirements for the degree of Master of Science in Industrial Psychology. The author wishes to express appreciation to Dr. Dannie J. Moffie under whose direction this study evolved.

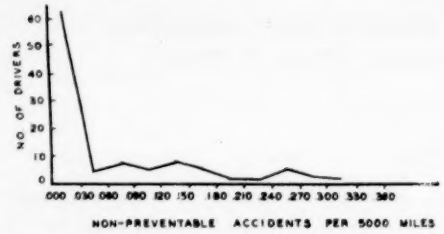


FIG. 2. Distribution of non-preventable accidents.

"Student's" *t*-ratio (2) was run between the means of the groups for each of the variables as follows: between the accident and non-accident groups, and between the upper and lower halves of the accident group.

The *F*-ratio test of differences was run between the variances of the groups mentioned above, and in like manner. This was done in order to determine if the distributions being compared by means of the *t*-ratio had significantly different dispersion within the distributions. In any analysis where the *F*-ratio was significant at the 5 per cent level of confidence or better, and the significance level of the *t*-ratio was critical, the significance levels of the *t*-ratios were corrected for the difference in variance (1).

On the basis of the analysis for preventable accidents per 5,000 miles between the accident and non-accident groups, those six variables having the most significant *t*-ratio were chosen to be analyzed by the Wherry-Doolittle Shrunken Multiple Correlation Technique (3). This type of analysis gives the maximum coefficient of correlation between the test variables and the criterion score after a correction has been made for the chance error added by each variable. The first step in this technique is the computation of intercorrelations among the variables included and the criterion (in this case, number of preventable accidents per 5,000 miles).

Results

Tables 1 and 2 show the condensed results of the study (only those variables showing significance out of the total of 43 included

Table 1

Variables Showing Significant Differences Between Upper Half, Accident Group and Lower Half, Accident Group

Variable	*t*	Difference in favor of
Preventable accidents		
Systolic Blood Pressure	2.397*	Lower half
Far Acuity, Right Eye	2.448*	Lower half
Far Acuity, Left Eye	2.231*	Lower half
Near Acuity, Left Eye	2.411*	Lower half
Kuder Mechanical Interest	2.268*	Lower half
MMPI Hypomania	2.226*	Lower half
Far Vertical Phoria	2.802**	Lower half
Near Acuity, Both Eyes	2.746**	Lower half
Kuder Artistic Interest	10.332***	Upper half
Kuder Literary Interest	5.181***	Upper half
Non-preventable accidents		
Marital Status	2.690*	Lower half
Far Acuity, Both Eyes	2.526*	Lower half
Near Acuity, Both Eyes	2.270*	Lower half
Near Acuity, Right Eye	2.456*	Lower half

* = $P < 0.05$.
** = $P < 0.01$.
*** = $P < 0.001$.

Table 2

Variables Showing Significant Differences Between Accident Group and Non-accident Group

Variable	t	Difference in favor of
Preventable accidents		
MMPI Hypochondriasis	2.011*	Non-accident group
MMPI Masculinity	2.136*	Non-accident group
Kuder Literary Interest	2.358*	Accident group
MMPI Psychasthenia	2.649**	Non-accident group
Kuder Artistic Interest	3.404***	Accident group
Non-preventable accidents		
Marital Status	2.413*	Non-accident group
Depth Perception	3.040**	Non-accident group

* = $P < 0.05$. ** = $P < 0.01$. *** = $P < 0.001$.

in the study are mentioned for the sake of brevity). In the tables, "difference in favor of" refers to which half of the sample distribution has the higher mean score. With respect to Marital Status and Systolic Blood Pressure, higher score means that the driver is married and has higher systolic blood pressure than the mean for the sample.

As a result of the Wherry-Doolittle Shrunken Multiple Correlation Analysis, it was found that only four of the six variables could be used to obtain the maximum multiple coefficient of correlation; these four were: Kuder Literary Interest, MMPI Hypochondriasis, Masculinity, and Schizophrenia. The resulting multiple coefficient of correlation was 0.36. This coefficient is statistically significant at the 1 per cent level of confidence. The resulting prediction equation is:

$$X_c = 0.00234834X_{35} - 0.00376978X_{42} - 0.00183060X_{39} + 0.00271910X_{31} + 0.14349491.$$

In this equation the symbols have the following meaning:

X_c = the predicted criterion score, number of preventable accidents per 5,000 miles
X_{35} = the MMPI Hypochondriasis score
X_{42} = the MMPI Schizophrenia score
X_{39} = the MMPI Masculinity score
X_{31} = the Kuder Literary Interest score

The standard error of a predicted criterion score is 0.1007.

Discussion

The analysis indicates that the following variables—Systolic Blood Pressure; Far Acuity, Right Eye; Far Acuity, Left Eye; Kuder Mechanical Interest; MMPI Hypomania; Far Vertical Phoria; Near Acuity, Both Eyes; Kuder Artistic Interest; and Kuder Literary Interest—are related to preventable accidents. That is, those drivers having high scores on all these variables, with the exception of Kuder Artistic and Literary Interests, tend to have a lower number of preventable accidents. Those drivers having high scores in Kuder Artistic and Literary Interests tend to have a higher number of preventable accidents.

In looking at the analysis of the non-preventable accident data, we see that the psychological test data variables have dropped out, and that only the following appear to be related to non-preventable accidents—Marital Status; Far Acuity, Both Eyes; Near Acuity, Both Eyes; Near Acuity, Right Eye; and Depth Perception. These all bear the same relationship to the non-preventable accident data; that is, the higher the score, the fewer the number of non-preventable accidents.

It might be possible to postulate from the results of this study that one of the main distinctions between preventable and non-pre-

ventable accidents is the fact that preventable accidents seem to be related to psychological test data. Similarly, it might be said that such sensory capacities as the visual skills and certain personal history data such as Marital Status seem to be related to non-preventable accidents.

On the basis of the Shrunken Multiple Correlation Analysis, it was shown that by using the four variables—Kuder Literary Interest, MMPI Hypochondriasis, Masculinity, and Schizophrenia—a reasonably satisfactory prediction of the number of preventable accidents per 5,000 miles could be made. This part of the study seems to be quite significant. It would suggest that the psychological test variables, particularly those dealing with personality traits, are the variables which contribute to the discovery of those drivers having a high preventable accident rate. It might be said that this would seem to bear out the results of the first part of the study.

It should be mentioned that considerable selection has already taken place before these particular drivers are hired by the company. The mere fact that they attended the North Carolina State College Driver Training School makes them a select or special population, and therefore unlike what might be called representative of the general population of commercial truck drivers. We do not have any "extreme" cases in the sample, and it is these who seem to contribute a great deal to any analysis, especially one in which an attempt is being made to isolate variables which are contributing to job success.

It must also be noted that the apparent significance of the *t*-ratios reported is inflated. This is due to the fact that those reported are selected from a total of 43 for each phase of the analysis of significance of differences.

Summary and Conclusions

This study has been concerned with studying the relationship of certain psychological and personal history data to the accident records of a sample of commercial truck drivers. The accident data used were of two types—preventable accidents and non-preventable accidents, and were equated for the number of miles driven.

The subjects for the study were 104 commercial truck drivers employed by a large East Coast trucking concern.

For each of the criterion groups two types of analyses were made: (a) the significance of differences between the upper and lower halves of the accident group was computed; and (b) the significance of differences between the accident and non-accident group was computed. In addition, a Wherry-Doolittle Shrunken Multiple Coefficient of Correlation was computed for four of the variables and the criterion score based on the first analysis of significance of differences.

From this study the following conclusions can be drawn:

1. A difference seems to exist between preventable and non-preventable accidents.

2. Psychological traits, as well as sensory capacities, are important in analyzing the accident liability for preventable accidents, while only personal history data and sensory capacities seem to be important in analyzing the accident liability for non-preventable accidents.

This study also demonstrated the applicability of a technique for controlling exposure to accident hazard; that is, using the number of accidents per unit mileage rather than just the total number of accidents.

Received January 30, 1953.
Early publication.

References

1. Edwards, A. L. *Experimental design in psychological research.* New York: Rinehart, 1950.
2. Guilford, J. P. *Fundamental statistics in psychology and education.* (2d Ed.) New York: McGraw-Hill, 1950.
3. Stead, W. H., Shartle, C. L., et al. *Occupational counseling techniques.* New York: American Book Company, 1940.

THE JOURNAL OF APPLIED PSYCHOLOGY
Vol. 37, No. 4, 1953

Applied Psychology in Action

A New Management Tool for Top Executives

The first issue of *Social Science Reporter* appeared on April 15, 1953, and is to be issued semi-monthly. It is an effort to build a bridge of better understanding between social scientists and the business leaders of America. The editor and publisher is Rex F. Harlow, 365 Guinda Street, Palo Alto, California, who wishes to be kept informed on the research that is being done. His staff will also seek to secure material for publication through personal interviews with scientists, the reading of scientific periodicals and attendance at scientific meetings.

The first issue contained abstracts of ten research projects such as: management organization and motivation (W. F. Whyte); management and the individual in an organization (E. Wight Bakke); multi-relational sociometric survey (I. R. Wechsler, R. Tannebaum, and E. Talbot); administering changes (Harriet O. Ronker and P. R. Lawrence); executive retirement (H. R. Hall); management-employee communication (Nat. Soc. of Prof. Engineers); improving supervisors (N. F. Maier); pension-plan policies and practices (M. Puchek); changing attitudes (L. Festinger and H. Kelley); mass persuasion (D. Cartwright).

Background of an Industrial Psychologist

Excerpts from a letter and enclosure of personal data may be of interest and value to all concerned with putting psychology to work in business and industry. After congratulating Dr. Marion A. Bills for her paper on "Our Expanding Responsibilities" and for developing an *Applied Psychology in Action* section in the April, 1953 issue of the *Journal of Applied Psychology*, Dr. John F. Michael wrote: "More than ten years ago I decided to become a psychologist with both feet solidly entrenched in industry. Consequently I have had a series of jobs each selected with the criterion being the extent to which it would contribute to a comprehensive industrial and business background. The two page enclosure, which might interest you, described this experience in greater detail.

"This attitude toward my professional relation to industry found great support in L. L. McQuitty's article in *American Psychologist* of several years ago. He also felt the great need of first being in demand to business not solely because of one's psychological background but mainly because of one's business background. The implications for the problem of communications are obvious. This is notwithstanding the fact that one of our largest consulting firms in industrial psychology in selecting their own staff maintains they do not particularly consider any possible business experience of the psychologists they hire.

"It seems fashionable to decry the small published output from psychologists spending all their time in industry. I fail to agree with the typical reasons offered by others for this. The main cause appears to me to be not any dire result of greater income, or any decrease in the desire to contribute to the general advance of the field, but merely the result of the lack of time and the need to sell to management the 'problem' of research. As a result it is almost impossible to do research as defined by other colleagues who have more time. An interesting support of this time factor might be the reading habits of the psychologist in industry. My own observation with reference to myself and others is that there is an inevitable decrease in professional reading from the dropping of the journal club subscriptions to the selection of a few journals none of which are completely read. I feel that such a decrease has not taken place so widely among our academic colleagues.

"The proposed feature is a most excellent suggestion. Comments to be accepted for publication should not be restricted too much with reference to content. Observations worthy of possible restatement into testable hypotheses, needs for specific techniques, notes as to application of psychological principles to unique problems in business are ex-

amples of possible contributions. In general, the basic characteristic should be brevity. This proposed section should give the psychologist in industry a voice by which to record some of his many varied actions which might be of interest and value to others. The thought of writing a comprehensive article likely has stifled the expression of many a good idea which, perhaps, would have required only several short sentences anyway. These ideas should not be lost.

"I fully realize nothing has been said to add any value or new points to Dr. Bills' speech. It is just a case of my being very pleased with her excellent discussion of the psychologist in industry with special reference to those of us who do not wear a tag which says 'psychologist.'"

Here are the high-lights of this industrial psychologist's background and training:

Age: 35; *Education:* AB, BS in LS, MA, and BBA, Western Reserve University; PhD, Ohio State University in 1952. Majors: Bibliographic research, general business administration, psychological aspects of marketing and merchandising.

Work History: (1) Business reference assistant, city public library, 3 years; (2) Research analyst, procurement division, Quartermaster Depot, 3 years; (3) Market analyst, for a large tire and rubber company, 3 years; (4) Jobs in The Lincoln Electric Company of Cleveland, Ohio, since January 1951: automatic saw operator, plant layout detailer, assistant purchasing agent, factory engineer, management analyst, and at present management engineer.

Professional Organizations: (1) American Psychological Association, Associate member since 1947; (2) American Marketing Association; (3) Cleveland Psychological Association; (4) Personnel Psychologists of Northern Ohio.

Noise and Absenteeism

A study reported by Dr. de Almeida is concerned with the maintenance of a work force of tabulating machine operators. He was called in because of absenteeism in a tabulating department of 125 workers. Noise was found to be a real source of trouble. He found that this could lead to aural lesions and neuropathic troubles. He recommended a rearrangement of equipment, placing machines on rubber supports, and "Cellotex" insulation of walls. These changes dropped the decibel average from 42 to 14. He states that 50 to 60 decibels, continued, causes permanent damage to the ear. Absences were reduced by 23 per cent, or 430 man/days in 6 months.—H. Ribeiro de Almeida, M.D. (Sao Paulo, Brazil). Influence of electric punch card machines on the human ear. *Archives of Otolaryngology*, 1950, **51**, 215–222.

Book Reviews

Heneman, H. G., Jr. and Turnbull, J. G. (Editors). *Personnel administration and labor relations: a book of readings.* New York: Prentice-Hall, Inc., 1952. Pp. xiv + 434. $3.95.

Pigors, P. and Myers, C. A. (Editors). *Readings in personnel administration.* New York: McGraw-Hill Book Company, Inc., 1952. Pp. xii + 483. $4.50.

Although not explicitly described as such, these two publications serve as supplementary reading sources for Yoder's *Personnel management and industrial relations* and Pigors and Myers' *Personnel administration,* respectively. They reflect the same similarities and differences as the basic texts themselves. Both attempt to present selected readings dealing with the various aspects of the field, the former stressing the "practical operating problems that confront the personnel man, labor relations man, or union leader in everyday life" and the latter "the philosophy of personnel administration, its basic problems and limitations, as well as criticisms and doubts raised by union leaders."

Heneman and Turnbull group their selections under four parts: The Setting of Industrial Relations; Personnel Administration; Labor Relations; and Research and Evaluation. They present 170 readings in all, many of them not being the entire article but excerpts selected to deal specifically with the topic under consideration. This method of presenting readings has certain advantages and disadvantages. It reduces overlap and duplication of content but at the same time requires a clear over-all structure to tie the excerpts together. This the authors have attempted to provide by including introductory statements before each part and by presenting a brief abstract, at the beginning of each chapter, of the main points of each reading and its contribution to the topic being covered. Despite the editors' care, some of the excerpts still "hang in mid-air" and the reader may feel a need for the omitted material which structured the original articles.

In contrast, Pigors and Myers present only 46 readings, organized under six parts: Nature and Scope of Personnel Administration; Analyzing and Handling Personnel Problems; The Foreman; Building and Maintaining Work Teams; Wage and Work Assignments; and Employee Services and Programs. The individual readings tend to be longer and to be more self-contained (than those in Heneman and Turnbull) and appear to be the complete original rather than excerpts. They also, however, are organized under sub-topics and each part includes an introductory statement, by the editors, which interprets the reading and shows its contribution to the topic being treated.

It would be impossible to rate one of these publications as better than the other; they are just different. Although both provide adequate coverage of the general field and have eight authors in common, there is no selection which appears in both. Each has its favorite sources—Heneman and Turnbull the University of Minnesota Industrial Relations Center, and Pigors and Myers the Harvard Business Review—but both draw upon American Management Association publications, the Personnel Journal, recent books, and related fields.

Both of these publications are worth being in the industrial psychologist's library. They have a double value—for what is in them and for what is not. The readings themselves are drawn from a wide variety of sources and disciplines and indicate clearly that the industrial psychologist, to be truly effective and to be able to communicate with management, must become conversant with a wealth of literature outside of psychological journals. That psychologists are becoming recognized as contributors to the field is indicated by the fact that slightly over 10% of the authors in each book are psychologists. However, these are the psychologists who have written articles for the personnel journals or who have participated in conferences run by personnel management groups. It seems clear that psychologists cannot be too modest and coyly wait for management to wade through psychological journals to find out what psychology has to offer. At any rate, only a very few of the readings in these two publications are taken from psychological journals.

It is also instructive to note in which areas authors from psychology are used. As one

might expect, Selection and Placement, Training, and Research are those in which the psychologist is recognized as making the most unique contribution, but other areas are beginning to be influenced, areas such as morale, safety, job analysis, and even administration. However, if these two publications are a guide, psychologists have still not made themselves felt as contributors to an understanding of labor-management relation, incentive plans, union activities, grievances, discipline, etc. Either we have nothing useful to offer on these problems or have not made it readily available to nonpsychologists. I hope it is the latter. Albert S. Thompson

Teachers College,
Columbia University

Walker, C. R. and Guest, R. H. *The Man on the assembly line.* Cambridge, Massachusetts: Harvard University Press, 1952. Pp. 175. $3.25.

The authors have contributed this pilot study in an effort to increase our general knowledge of the adjustments made and the satisfactions derived by workers on an assembly line.

It is a thorough, well-organized inquiry into the expressed feelings and attitudes of these men regarding many facets of their work. The book describes in detail the work climate at Plant X which is an automobile assembly plant. The method of investigation was both qualitative and quantitative, giving a clear picture of attitudes toward seven characteristics of the work. These, as listed by the authors, were: "the worker's immediate job, his relations to fellow workers, pay and security, his relation to supervision, general working conditions in the plant, promotion and transfer, and his relation to the union." These attitudes were determined by interviews, questionnaires, and the study of overt behavior.

The results of this investigation are not new or startling. It is a typical study of attitudes of employees toward their work situation, using the usual attitude research methodology showing us which aspects of the work were most liked and disliked.

Unfortunately, the authors do not emphasize the main reason the employees *liked* their work. "Pay." This criticism can be levelled at many other social scientists of course. Perhaps it is because this aspect is considered out of their realm and they can't do much about it. On the other hand, they have come up with some realistic suggestions for alleviating, at least in part, the main *dislikes* which these men expressed concerning their work which was the immediate job content. This further inquiry into the specific aspects of the job content itself which the men objected to is the only unique contribution of the book. Evidently, it was not hard work itself but the mechanical pacing and the repetitiveness of the job to which they objected most strenuously. Walker and Guest have proposed what appear to be practical suggestions for minimizing these objections.

The essence of the problem seems to be that we have progressed too rapidly in the technique of mass production and it is time now to stop and re-organize our thinking to take into consideration the worker himself in order to solve some of the problems created. Many industrial engineers and members of plant management, however, are going to disagree until we can prove its economic effectiveness. John M. Cook

Radio Corporation of America,
RCA Victor Division, Camden, N. J.

Laird, D. A. and Laird, Eleanor C. *Practical sales psychology.* New York: McGraw-Hill, 1952. Pp. xii, 291. $4.00.

The formula adopted by the authors of this book is clear and easy to follow up to a point. Take self-administering tests of twelve to eighteen items each and label them self-sufficiency, dominance, self-confidence, social mixing, friendliness, hygienic habits, fatigue, tactfulness, kindness, sympathy, warm-heartedness, optimism, ability to meet emergencies, considerateness, snobbishness, egotism, fault-finding, self-centeredness, and hot-temperedness. This will show the reader "where you stand in some qualities which are important for the salesperson" (p. 117). It will also help to convince him that scientific psychology underlies your approach.

Sprinkle with authoritative statements such as: "About one person out of four has some characteristics which make him disagreeable" (p. 168). "Records show that graduates who received poor grades in college become as good

salesmen as those who had high grades" (p. 12). "Psychological studies have shown that the love of selling or 'sales drive' is essential for sales success" (p. 13). "Ninety-eight per cent of experienced sales people like selling better than any job" (p. 15).

Prove your points with "studies." For example, if you want to demonstrate that love of selling can be developed, quote a study which shows that the longer men have been selling fire protection equipment the better they like it. Do not mention the possibility that self-selection rather than development accounts for the results.

Lard with references to big names and stories of the Horatio Alger type. Mention N. W. Ayer, Marshall Field, John Wanamaker, Owen D. Young, Abraham Lincoln and Colonel Edward M. House. Tell how they succeeded. In the case of Colonel House quote him as saying, "Let the other fellow make the mistake first." Point out that "House gained political power in the Woodrow Wilson administration by following that policy" (p. 264).

Do not allow difficulties to disturb you. If the concept of the salesman implies a generality which does not exist, if what is true of selling soap is not necessarily true of selling bonds, say so and then forget it. If the criterion of success in selling is so difficult to establish that it represents a major problem to psychologists in that field, don't even say so. Ignore it.

On the other hand, abash reviewers such as this one by writing with a style and clarity that far exceeds what your more conservative brethren produce. Demonstrate that you have absorbed the best of the common-sense stereotypes about the sales situation and that you have a rare gift for organization and emphasis. Beat all of the "Sales Power in Five Easy Lessons" boys at their own game. Convince even the psychologists that, after all, salesmen are probably better off to read your book than most of the literature that is now given them.

But do not say that "The first attempts to teach something about human nature in relation to the customer produced a lot of plausible-sounding nonsense. Magazines and books were filled with 'how to sell' advice which made psychologists, who knew human nature, laugh aloud" (p. 38). *Who's laughing?*

S. Rains Wallace, Jr.
Life Insurance Agency Management Association, Hartford, Connecticut

Barlow, F. *Mental prodigies.* New York: Philosophical Library, 1952. Pp. 256. $4.75.

This book is described on its title page as "An enquiry into the faculties of arithmetical, chess and musical prodigies, famous memorizers, precocious children and the like, with numerous examples of 'lightning' calculators and mental magic." Actually most of it is devoted to lightning calculators and to examples of the mathematical tricks, stunts and problem solving used in performances of "mental magic." The author is himself something of a "mathemagician," having given performances using his own mnemonic devices and calculating short-cuts. His interest in such matters dates back to the nineteen-twenties or earlier and has led him to interview several arithmetical prodigies. For the most part, however, his knowledge of prodigies is second or third hand.

The first chapter, 58 pages in length, is an interesting account of nineteen outstanding arithmetical prodigies of the last two and a half centuries, together with brief notes on twenty-two others less well known. The large majority of both groups were born between 1700 and 1870, and only seven of them since 1900. For the outstanding cases the author gives where possible date of birth and of death, nationality, age when the special gift became evident, amount of education, examples of calculating feats performed, and an estimate of general ability. Here he draws heavily from Scripture's 1891 article on "Arithmetical prodigies" in the *American Journal of Psychology,* and from Mitchell's 1907 article on "Mathematical prodigies" in the same Journal. As most of the cases reported upon antedated scientific psychology, their histories, with a few notable exceptions, are sketchy and poorly documented. The author cites uncritically hear-say evidence, articles from old newspapers and popular magazines, and writers on psychic research. He also has a few systematic biases, including: (1) the usual bias of one who has dabbled in psychic research; (2) an admitted "instinc-

tive dislike of precocious children"; (3) a tendency to exaggerate the one-sidedness of prodigies and to underestimate their general ability. The third of these may be an outcome of the second. For example, the Belgian boy Verhaeghe is described as "an adolescent of seventeen with the mental age of a babe of two years." Yet this boy, when examined by a committee of mathematicians in 1946, gave the fourth power of 1,246 in 10 seconds, the sixth root of 24,137,585 in 25 seconds, and the square of 888,888,888,888,888 in 40 seconds!

The arithmetical prodigies whose nationalities are given distribute as follows by country: Britain 10, France and Italy 7 each, United States 5, Germany 2, and 1 each for India, Ceylon, Greece, Belgium, Spain, Switzerland, Egypt, and Mexico. Britain's lead is no doubt due to the fact that the author has covered his own country more thoroughly than other parts of the world. Of the United States' five, two were Negro slaves (one of them born in Africa), and two were from the little state of Vermont. The author's list includes two very famous scientists (Gauss and Ampère), a father-son pair (George P. Bidder, Sr. and Jr.), and six university graduates (both Bidders, Gauss, Ampère, the American Safford who became a professor of astronomy, and Whately, an Archbishop of Dublin). Several others gave evidence of superior general intelligence. At the opposite extreme three were considered mental defectives, though the basis for such classification is dubious in two of the three cases. Several described as dull were almost certainly well above average in general ability. Of the 41 arithmetical prodigies, four were sheep herders in childhood, two were blind, two had 12 fingers and 12 toes, and one was born without legs or arms. The relatively high incidence of sheep herding and of physical anomalies is hardly explicable in terms of chance, nor is the sex ratio of 39 males to 2 females.

The book would have been more acceptable to psychologists if it had included only the chapters on History and Data (58 pages), The Calculations Considered (15 pages), Development (6 pages), Famous Memorizers (19 pages), Mnemonics (9 pages) and Mental Magic (53 pages). The remaining hundred pages are a hodgepodge of naive comment on such topics as heredity and instinct, mental imagery, chess and musical prodigies, precocity and genius, and the subconscious.

The book is entertainingly written. The authenticated feats of calculation and memory which the author has recounted are fascinating and astonishing. The chapters on mental magic will intrigue many readers by the revelations given about such things as naming the day of the week on which a dated event occurred or will occur; short-cut methods for a great variety of computations, including among others squaring, cubing, root extraction, translating months or years into seconds and miles into inches or barleycorns; and directions for carrying out such parlor stunts as "think of a number," "draw any card," "a loan and a present," "change for a shilling," and numerous others. By showing that many feats which appear so difficult can be mastered by almost anyone with a moderately high IQ and a fair memory, these chapters can be expected to increase the number of amateur mathemagicians whose performances will probably amuse, mystify, and bore us in about equal proportions.

Lewis M. Terman

Stanford University

Dunsmoor, C. and Davis, O. *How to choose that college.* Boston: Bellman Publishing Co., 1951. Pp. 51. $.90 (cloth bound).

The rapidly expanding number of students considering college training makes the need for materials to assist them with their choices increasingly urgent. This book was written for high school students and their parents to discuss many of the questions which they typically have. It is simply written and attractively illustrated. The authors systematically consider, among other things, such problems as college requirements, application, financial plans, and making good once admitted. This is quite an order for fifty-one pages.

To cover this ground, the authors rely heavily on lists of things to do and avoid, on typical recommended high school curricula, considerable oversimplification, and frequent suggestions to see a counselor. The book would, therefore, seem to be best suited for

use in a guidance class; without expert fill-in, most students and their parents would probably either be left with unanswered questions or be misled.

Counseling plays an interesting dual role in this book. On the one hand, the student is left in rather large measure to decide for himself whether he has the right goals, interests, and abilities. This is in spite of the ample available evidence regarding the tendency to appropriate to the self desirable characteristics in such circumstances. On the other hand, the student is sent to the counselor for such things as assistance with writing of letters of application and, immediately on receipt of acceptances, for "review" of these. This suggests a dependency likely to plague college officials and counselors later on.

Properly handled, this book should make a useful contribution to the literature on guidance. But, because it has been almost oversimplified and condensed, it is doubtful whether it should be turned over to students as the sole and sufficient guide through a difficult terrain.

John W. Gustad
University of Maryland

Argyris, C. ***An introduction to field theory and interaction theory.*** New Haven: Yale Labor and Management Center, 1952. Pp. 71. $1.00.

This brief 71-page pamphlet was originally prepared for members of the Labor and Management Center at Yale University. Presented in each case are some basic foundations of each theory, some assumptions underlying each theory, some of the goals of each theory, some basic operations suggested by each theory, and some of the more important concepts utilized in each theory. The writing is clear in so far as the somewhat turgid language of the authors, whose work is summarized, makes that possible, and simple in so far as the extreme complexity of the theories discussed allows. As a presentation of field theory and interaction theory, therefore, this book may be recommended.

There are several points, however, on which the critical reader may well feel uneasy. In the first place, the author makes no attempt to link up the two sets of theories he is dealing with. Is field theory compatible with interaction theory? Do they deal with the same aspects of human behavior? Are there alternative descriptions of similar phenomena? Is there any evidence of an experimental kind to make a choice between them possible? No information is given on these vitally important points.

Again, in what way do the concepts and ideas of Lewin, on the one hand, and of Chapple, Arensberg, Whyte, and Homans, on the other hand, constitute a theory? In science the term theory is ordinarily used to denote a system of hypothetical constructs or intervening variables which enables us: (a) to summarize existing knowledge; and (b) to make deductions leading to testable predictions. This clearly is not the purpose of either field or interaction theory. Both are essentially semantic attempts, occasionally leavened by the use of unorthodox mathematics, to persuade the reader to accept unproven views and to make use of complex ways of stating what often appear to be obvious, commonsense notions. Argyris never comes to grips with the problem of proof and disproof of theories of this kind, nor does he attempt to show that they have a function to perform which could not equally well be performed in other ways.

This failure to provide any form of critical discussion is the major weakness of the book. The unwary reader might think from Argyris' presentation that Lewin's use of topological concepts, or hodological space, is acceptable to orthodox mathematicians. This, as far as I know, is not so, and one might have expected at least a brief reference to criticisms of these, as of many other points. A book of this kind will not be very useful to the initiated, who will be familiar with its contents in any case; nor will it help the uninitiated to obtain a balanced view of the theories proposed. The habit of stating a position without answering important criticisms of that position, or even mentioning that such criticisms exist, is unfortunately widespread in psychology. This book is an outstanding example of what, to the reviewer, appears to be a very bad practice indeed.

H. J. Eysenck
Psychology Department, Maudsley Hospital
London, England

New Books, Monographs, and Pamphlets

Books, monographs, and pamphlets for listing and possible review should be sent to Donald G. Paterson, Editor, Department of Psychology, University of Minnesota, Minneapolis 14, Minnesota

Introduction to the Rorschach technique. Robert M. Allen. New York: International Universities Press, Inc., 1953. Pp. 126. $3.00.

Introduction to exceptional children. Revised edition. Harry J. Baker. New York: Macmillan, 1953. Pp. 500. $5.00.

Social psychology. Hubert Bonner. New York: American Book Co., 1953. Pp. 439. $4.25.

Progress in clinical psychology. Vol. I, Sec. 1. Daniel Brower and Lawrence E. Abt, Editors. New York: Grune and Stratton, 1952. Pp. 328. $5.75.

The individual and world society. P. E. Corbett. Princeton: Center for Research on World Political Institutions, Princeton University, 1953. Pp. 59. Gratis.

The MMPI: a review. William C. Cottle. Lawrence, Kansas: School of Education, Univer. of Kansas Publ., 1953. Pp. 82.

Logic and language. Second series. A. G. N. Flew, Editor. New York: Philosophical Library, 1953. Pp. 242. $4.75.

The Grassi Block Substitution Test for measuring organic brain pathology. Joseph R. Grassi. Springfield: Charles C Thomas, Publisher, 1953. Pp. 75. $3.00.

Initiating and administering guidance services. S. A. Hamrin. Bloomington, Illinois: McKnight and McKnight, 1953. Pp. 220. $3.00.

Introduction to statistical methods. Palmer O. Johnson and Robert W. B. Jackson. New York: Prentice-Hall, 1953. Pp. 394.

Heredity in health and mental disorder. Franz Josef Kallmann. New York: W. W. Norton Co., Inc., 1953. $5.00.

Problem drinkers can be helped. G. N. Lansdown. Devon, England: Arthur H. Stockwell Limited, 1953. Pp. 72.

Psychology of industrial relations. C. H. Lawshe. New York: McGraw-Hill Book Co., Inc., 1953. Pp. 350. $5.50.

Comparative conditioned neuroses. Roy Waldo Miner, Editor. New York: New York Acad. of Sci., 1953. Pp. 379. $3.50.

The natural superiority of women. Ashley Montagu. New York: The Macmillan Co., 1953. Pp. 205. $3.50.

God, labor and management. Alfred Morgan. New York: The William-Frederick Press, 1953. Pp. 28. $1.00.

Group psychotherapy. Florence B. Powdermaker and Jerome D. Frank. Cambridge: Harvard Univer. Press, 1953. Pp. 615. $6.50.

The conception of disease. Walther Riese. New York: Philosophical Library, 1953. Pp. 120. $3.75.

The universe of meaning. Samuel Reiss. New York: Philosophical Library, 1953. Pp. 227. $3.75.

Philosophy and the ideological conflict. Charles S. Seely. New York: Philosophical Library, 1953. Pp. 319. $5.00.

The retention of meaningful material. Joseph Francis Sharpe. Washington, D. C.: Catholic University of America Press, 1952. Pp. 66. $1.00.

Capitalism overhauled. Job Socius. New York: The William-Frederick Press, 1952. Pp. 79. $2.00.

Introduction to testing and the use of test results in public schools. Arthur E. Traxler, Robert Jacobs, Margaret Selover and Agatha Townsend. New York: Harper and Brothers, 1953. Pp. 113. $2.50.

Retirement and the industrial worker. Jacob Tuckman and Irving Lorge. New York: Bureau of Publications, Teachers College, Columbia Univer., 1953. Pp. 105. $2.75.

The science of color. Committee on Colorimetry of the Optical Society of America. New York: Thomas Y. Crowell Co., 1953. Pp. 385. $7.00.

Evaluating research and development. Irving R. Weschler and Paula Brown, Editors. Los Angeles: Institute of Industrial Relations, University of California, 1953. Pp. 104. $1.65.

Research Report of U. S. Naval School of Aviation Medicine. Pensacola: U. S. Naval Air Station, 1953.

REVUE DE PSYCHOLOGIE APPLIQUÉE

PUBLICATION TRIMESTRIELLE

Directeurs : Dr P. PICHOT et P. RENNES

TOME 3 - NUMÉRO 3 - JUILLET 1953

NUMÉRO SPÉCIAL

consacré au

COMPTE-RENDU DE LA SECTION DE PSYCHOLOGIE CLINIQUE DU CONGRÈS INTERNATIONAL DE PSYCHOTECHNIQUE

(Paris, 27 Juillet - 1er Août 1953)

THÈMES PRINCIPAUX

— Rapports déontologiques entre psychologues et psychiatres.

— Psychologie de l'épilepsie.

Texte intégral (français ou anglais) des rapports et communications

English summaries of the French texts - Résumés français des textes anglais

---

Prix de ce numéro : France : 600 fr - Étranger : 800 fr — Les abonnés recevront ce numéro sans majoration — Abonnements 1 an : France : 1000 fr - Étranger : 1300 fr

Rédaction et Administration : 15, Rue Henri-Heine - PARIS 16e — C. C. P. Paris 5551-62

Prentice-Hall
COLLEGE TEXTS

APPLIED PSYCHOLOGY

By HAROLD E. BURTT, *Ohio State University*

- A highly readable text which combines experimental and clinical approaches to describe in intelligible style the contributions of psychology to American culture. Called an exceptionally adequate description of psychology at work in its currently most important fields, the treatment embodies an organization and analysis of all major research. It is distinctive in integrating various fields as well as in covering particular areas.

821 pages · 6" x 9" · 1948

APPLIED PSYCHOLOGY—Abridged Edition

By HAROLD E. BURTT, *Professor of Psychology and Chairman of Department of Psychology, Ohio State University*

- Shortened by half and accelerated in readability, this version of Burtt's APPLIED PSYCHOLOGY (see above) is designed for an average college course. It adequately covers applications in education, medicine, law and business—omits marginal material which is treated comprehensively in the original.
- The Abridged Edition shows the student what psychologists are doing in a practical way in public schools, clinics, social agencies, juvenile courts, employment agencies, factories, and business offices.

465 pages · 5½" x 8½" · July 1952

COUNSELING AND PSYCHOLOGY: Vocational Psychology and Its Relation to Educational and Personal Counseling

By MILTON L. BLUM and BENJAMIN BALINSKY, *College of the City of New York*

- This eclectic treatment combines major viewpoints on counseling and guidance in a systematic approach.
- Explains all aspects of the guidance process and treats theory in direct relation to applications. Illustrative cases are drawn from the authors' own files.
- Coordinates basic psychological principles with such tools of guidance as vocational, educational and personal counseling, vocational tests, interest classifications, and occupational information.

586 pages · 5½" x 8½" · Published 1951

THE COUNSELING INTERVIEW

By CLIFFORD E. ERICKSON, *Dean of Michigan State College*

- A practical, common-sense text on ways to develop skill in interviewing, this text examines all aspects of the interview situation, makes concrete suggestions, and provides case materials for class discussion.

For approval copies write

PRENTICE-HALL, Inc. 70 FIFTH AVENUE, NEW YORK 11, N.Y.